

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2002, 00:12:53 : Search time 1474.82 Seconds
(without alignments) updates/sec
5190.255 Million cell

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464
Sequence: 1 gaagagagaagaatgga.....ttggcaaacgacgagac 464

Scoring table:

OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.om:*
20: em.ov:*
21: em.pat:*
22: em.ph:*
23: em.pl:*
24: em.ro:*
25: em.sts:*
26: em.un:*
27: em.vi:*
28: em.vi:*
29: em.vi:*
30: em.htgo_hum:*
31: em.htgo_inv:*
32: em.htgo_rnd:*
33: em.htg_hum:*
34: em.htg_inv:*
35: em.htg_rnd:*
36: em.htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	464	100.0	1940	10 AF351130	AF351130 Rattus no
2	40	8.6	2088	10 CRGAD34X	L28147 Hamster Gad
3	35	7.5	2275	10 MMMPRMR	X51829 Mouse myelo
4	35	7.5	215734	2 AC073828	AC073828 Mus muscu
5	23	5.0	216	10 MMU83984	U83984 Mus muscu
6	23	5.0	224	10 CGU83983	U83983 Cricetus
7	22	4.7	185414	2 AL355526	AL355526 Homo sapi
8	22	4.7	190856	2 AC024505	AC024505 Homo sapi
9	21	4.5	61147	2 AL162294	AL162294 Human DNA
10	21	4.5	155190	2 AC026323	AC026323 Homo sapi
11	21	4.5	165228	2 AC005325	AC005325 Homo sapi
12	20	4.3	1776	8 MTLPMPT95	Z50200 L. perenne m
13	20	4.3	2805	4 AF286474	AF286474 Bos tauru
14	20	4.3	108893	9 HSDJ130E4	AL078582 Human DNA
15	20	4.3	110000	2 PRMALAP1.0	AL034557 Plasmodiu
16	20	4.3	159644	9 AP000820	AP000820 Homo sapi
17	20	4.3	163186	9 AB045360	AB045360 Homo sapi
18	20	4.3	165079	2 AC024517	AC024517 Homo sapi
19	20	4.3	167300	2 AC021394	AC021394 Homo sapi
20	20	4.3	171326	2 AL353894	AL353894 Homo sapi
21	20	4.3	175504	2 AL591916	AL591916 Homo sapi
22	20	4.3	180815	2 AC025549	AC025549 Homo sapi
23	20	4.3	197926	2 AC073844	AC073844 Homo sapi
24	20	4.3	300000	9 AP002533	AP002533 Homo sapi
25	19	4.1	371	11 HSC009YH5	Z67713 H. sapiens D
26	19	4.1	8372	14 AF017780	AF017780 Sour cher
27	19	4.1	72928	6 AR103481	AR103481 Sequence
28	19	4.1	82918	8 AB010071	AB010071 Arbidops
29	19	4.1	97142	2 AC010439	AC010439 Homo sapi
30	19	4.1	114022	9 AL157952	AL157952 Human DNA
31	19	4.1	115391	2 AC022439	AC022439 Homo sapi
32	19	4.1	126936	2 AC090472	AC090472 Homo sapi
33	19	4.1	126936	2 AC090730	AC090730 Homo sapi
34	19	4.1	151037	9 AC018741	AC018741 Homo sapi
35	19	4.1	151673	10 AC087102	AC087102 Rattus no
36	19	4.1	151970	2 AC040905	AC040905 Homo sapi
37	19	4.1	162030	2 AC018601	AC018601 Homo sapi
38	19	4.1	167859	2 AC016391	AC016391 Homo sapi
39	19	4.1	171850	2 AC021405	AC021405 Homo sapi
40	19	4.1	178671	2 AC069283	AC069283 Homo sapi
41	19	4.1	186448	2 AC034129	AC034129 Homo sapi
42	19	4.1	190174	2 AC022739	AC022739 Homo sapi
43	19	4.1	191756	2 AL512785	AL512785 Homo sapi
44	19	4.1	202851	2 AC015684	AC015684 Homo sapi
45	19	4.1	206454	2 AC011227	AC011227 Homo sapi

ALIGNMENTS

RESULT 1
AF351130 LOCUS 1940 bp DNA ROD 09-APR-2001
DEFINITION Rattus norvegicus progression elevated gene-3 protein (Peg-3) gene,
AF351130 Promoter region.
ACCESSION AF351130 GI:13562027
VERSION AF351130.1
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1940)
REFERENCE Su,Z., Shi,Y. and Fisher,P.B.
AUTHORS Cooperation between AP1 and PEA3 sites within the progression
TITLE elevated gene-3 (Peg-3) promoter regulate basal and differential
JOURNAL expression of Peg-3 during progression of the oncogenic phenotype
Oncogene 19 (30), 3411-3421 (2000)


```
Medline 90251472 Location/Qualifiers
FEATURES
Source
  1..2275
    /organism="Mus musculus"
    /strain="SL"
    /db_xref="taxon:10090"
    /clone_lib="lambda ZAP"
    /clone="116"
    229..2202
      /note="Myd116 protein (AA 1-657)"
      /codon_start=1
      /protein_id="CA36128.1"
      /db_xref="GI:53041"
      /db_xref="SWISS-PROT:P17564"
      /translation="MAPSPRPOHVLHMRANHYLLSLPLGLLSRAMSLRCPPEPEA
      WLAKVTGADQIEAALLPPTVPSGNLLPHGETEESGSPESQAAQRLCLVAEESP
      ETWGLSNVDEYNAPGQDDLREKEMERTAGKATLQAGLQAGADKRGVEVAREGVAE
      PAYPTSOLEGGPAENEDGETVYQASASIAIPGKPSPTPVPLGEAEHQATEEGT
      ENKADPSNPSGSHRAMVYRSREKPKQEGAKVFAHRAAGGHPCRMAAEQGPET
      TPCVCTNATLKAHYRPGEDTEEDNSDSDSAEDTAQTGATPHTSAFLKAWYRPGEDTEEDNSDSDSAEDTAQTG
      DVEEDSDSDSAEDTAQTGATPHTSAFLKAWYRPGEDTEEDNSDSDSAEDTAQTG
      APHTSAFLKAWYRPGEDTEEDNSDSDSAEDTAQTGATPHTSAFLKAWYRPGEDT
      EDTEEDNSDSDSAEDTAQTGATPHTSAFLKAWYRPGEDTEEDNSDSDSAEDTAQTG
      EKDESPMAAPKLPRLRLRFLKAPTRQDPEIPLKARKVHAKEVTVFLAVMAGP
      AQAARRGPMQGFARDRSRFARIRIAQAEKLGPLYLTPDSRARAMARLRNPSLQSEPRS
      SSEATPLTDVTPSPPLPSTPSPSLXIGRRG"
      2275
    /note="polya site"
  polyA-site
BASE COUNT 577 a 652 c 636 g 410 t
ORIGIN
Query Match 7.5% Score 35; DB 10; Length 2275;
Best Local Similarity 100.0%; Pred. No. 1e-07; Mismatches 0; Gaps 0;
Matches 35; Conservative 0; Indels 0;

Qy 373 cgtcccgagcccgctgagggcgacatgaac 407
|||||
Db 96 CGTCCCGAGCCCGCCTGAGGGCGACATGAAC 130

RESULT 4
AC073828 215734 bp DNA HTG 29-JUN-2000
LOCUS Mus musculus clone Rp23-9j18, WORKING DRAFT SEQUENCE, 21 unordered
DEFINITION
AC073828
AC073828.1 GI:8810445
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
1 (bases 1 to 215734)
DOE Joint Genome Institute.
TITLE
Sequencing of Mouse
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 215734)
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1740562
Center clone name: RPI-23_9j18
-----
Summary Statistics
Consensus quality: 19667 bases at least Q40
```

```
Consensus quality: 207971 bases at least Q30
Consensus quality: 209928 bases at least Q20
Estimated insert size: 207740; agarose-fp estimation
Estimated insert size: 213734; sum-of-contrigs estimation
Quality coverage: 6.12 in Q20 bases; agarose-fp estimation
Quality coverage: 5.94 in Q20 bases; sum-of-contrigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1013: contrig of 1013 bp in length
* 1014 1113: gap of unknown length
* 1114 2175: contrig of 1062 bp in length
* 2176 2275: gap of unknown length
* 2276 3429: contrig of 1154 bp in length
* 3430 3529: gap of unknown length
* 3530 4884: contrig of 1355 bp in length
* 4885 4984: gap of unknown length
* 4985 6848: contrig of 1864 bp in length
* 6849 6948: gap of unknown length
* 6949 9317: contrig of 2369 bp in length
* 9318 9417: gap of unknown length
* 9418 12128: contrig of 2711 bp in length
* 12129 12228: gap of unknown length
* 12229 14570: contrig of 2742 bp in length
* 14571 15070: gap of unknown length
* 15071 19571: contrig of 4501 bp in length
* 19572 25961: contrig of 6290 bp in length
* 25962 26061: gap of unknown length
* 26062 36753: contrig of 10692 bp in length
* 36754 36853: gap of unknown length
* 36854 49116: contrig of 12263 bp in length
* 49117 49216: gap of unknown length
* 49217 61879: contrig of 12663 bp in length
* 61880 61979: gap of unknown length
* 61980 72513: contrig of 10534 bp in length
* 72514 72613: gap of unknown length
* 72614 83721: contrig of 11108 bp in length
* 83722 83821: gap of unknown length
* 83822 98633: contrig of 14812 bp in length
* 98634 98733: gap of unknown length
* 98734 118323: contrig of 19590 bp in length
* 118324 118423: gap of unknown length
* 118424 137668: contrig of 19245 bp in length
* 137669 137768: gap of unknown length
* 137769 162338: contrig of 24570 bp in length
* 162339 162438: gap of unknown length
* 162439 186625: contrig of 24187 bp in length
* 186626 186725: gap of unknown length
* 186726 215734: contrig of 29009 bp in length.

FEATURES
Source
  1..215734
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="Rp23-9j18"
    /clone_lib="RPI mouse BAC library 23"
BASE COUNT 54325 a 53055 c 52767 g 53583 t 2004 others
ORIGIN
Query Match 7.5% Score 35; DB 2; Length 215734;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 373 cgtcccgagcccgctgagggcgacatgaac 407
|||||
Db 8415 CGTCCCGAGCCCGCCTGAGGGCGACATGAAC 8449
```

```

RESULT 5
LOCUS MMU83984 216 bp DNA ROD 07-JUL-1998
DEFINITION Mus musculus apoptosis associated protein (GADD34) gene, promoter
sequence.
ACCESSION U83984
VERSION U83984.1 GI:3258621
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 216)
JOURNAL Hollander, M.C., Zhan, O., Bae, I. and Fornace, A.J. Jr.
Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
MEDLINE J Biol. Chem. 272 (21), 13731-13737 (1997)
97298078
REFERENCE
AUTHORS 2 (bases 1 to 216)
Hollander, M.C. and Fornace, A.J. Jr.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD
20892, USA
FEATURES
Source Location/Qualifiers
gene 1..216
/organism="Mus musculus"
/db_xref="taxon:10090"
<1..>216
/gene="GADD34"
/note="growth arrest and DNA damage-inducible gene 34;"
Myd116 gene
<1..>216
/gene="GADD34"
135..146
/gene="GADD34"
149..158
/gene="GADD34"
/feature="ATF/CRE site"
TATA_signal 177..182
/gene="GADD34"
BASE COUNT 40 a 66 c 63 g 47 t
ORIGIN
Query Match 5.0%; Score 23; DB 10; Length 216;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 203 ccgcctccggtgacgtacga 225
|||||
Db 138 cccgcctccggtgacgtacga 160
RESULT 6
LOCUS CGU83983 224 bp DNA ROD 07-JUL-1998
DEFINITION Cricetus griseus apoptosis associated protein (GADD34) gene,
promoter sequence.
ACCESSION U83983
VERSION U83983.1 GI:3258620
KEYWORDS
SOURCE Chinese hamster.
ORGANISM Cricetus griseus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus.
TITLE 1 (bases 1 to 224)
JOURNAL Hollander, M.C., Zhan, O., Bae, I. and Fornace, A.J. Jr.
Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
MEDLINE J Biol. Chem. 272 (21), 13731-13737 (1997)
97298078
REFERENCE
AUTHORS 2 (bases 1 to 224)
Hollander, M.C. and Fornace, A.J. Jr.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD
20892, USA
FEATURES
Source Location/Qualifiers
gene 1..224
/organism="Cricetus griseus"
/db_xref="taxon:10029"
<1..>224
/gene="GADD34"
/feature="growth arrest and DNA damage-inducible gene 34;"
<1..>224
/gene="GADD34"
135..146
/gene="GADD34"
149..158
/gene="GADD34"
/feature="ATF/CRE site"
TATA_signal 174..179
/gene="GADD34"
BASE COUNT 46 a 67 c 63 g 48 t
ORIGIN
Query Match 5.0%; Score 23; DB 10; Length 224;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 203 ccgcctccggtgacgtacga 225
|||||
Db 138 cccgcctccggtgacgtacga 160

```

```

RESULT 7
LOCUS AL355526 185414 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-152L7, *** SEQUENCING IN
PROGRESS ***. 22 unordered pieces.
ACCESSION AL355526
VERSION AL355526.4 GI:9797321
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 185414)
JOURNAL Plumb, B.
Direct Submission
Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9231038.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Project Information
Center project name: ba15217
----- Summary statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173494 bases at least Q40
Consensus quality: 177549 bases at least Q30
Consensus quality: 180091 bases at least Q20
Insert size: 183314; sum-of-contigs
Insert size: 188970; 9.8% error; agarose-fp
Quality coverage: 4.22x in Q20 bases; sum-of-contigs quality
coverage: 4.28x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 3619: contig of 3619 bp in length
* 3620 3719: gap of 100 bp
* 3720 8736: contig of 5017 bp in length
* 8737 8836: gap of 100 bp
* 8837 13069: contig of 4233 bp in length
* 13070 13169: gap of 100 bp
* 13170 42396: contig of 29227 bp in length
* 42397 42496: gap of 100 bp
* 42497 113471: contig of 70975 bp in length
* 113472 113571: gap of 100 bp
* 113572 119459: contig of 5888 bp in length
* 119460 119559: gap of 100 bp
* 119560 122149: contig of 2590 bp in length
* 122150 122249: gap of 100 bp
* 122250 125172: contig of 2923 bp in length
* 125173 125272: gap of 100 bp
* 125273 129826: contig of 4554 bp in length
* 129827 129926: gap of 100 bp
* 129927 133355: contig of 3429 bp in length
* 133356 133455: gap of 100 bp
* 133456 136798: contig of 3343 bp in length
* 136799 136898: gap of 100 bp
* 136899 140265: contig of 3367 bp in length
* 140266 140365: gap of 100 bp
* 140366 142977: contig of 2612 bp in length
* 142978 143077: gap of 100 bp
* 143078 146062: contig of 2985 bp in length
* 146063 146162: gap of 100 bp
* 146163 149355: contig of 3193 bp in length
* 149356 149455: gap of 100 bp
* 149456 154688: contig of 5233 bp in length
* 154689 154788: gap of 100 bp
* 154789 162417: contig of 7629 bp in length
* 162418 162517: gap of 100 bp
* 162518 170626: contig of 8109 bp in length
* 170627 170726: gap of 100 bp
* 170727 175237: contig of 4511 bp in length
* 175238 175337: gap of 100 bp
* 175338 177559: contig of 2222 bp in length
* 177560 177659: gap of 100 bp
* 177660 180444: contig of 2785 bp in length
* 180445 180544: gap of 100 bp
* 180545 185414: contig of 4870 bp in length.

```

FEATURES

source

```

1. 185414
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="1"
  /clone="RP11-152L7"
  /clone_lib="RPC1-11.1"
1. 3619
  /note="assembly_fragment:01810
  fragment_chain:1"
misc_feature
  3720..8736
  /note="assembly_fragment:01729
  fragment_chain:1"
misc_feature
  8837..13069
  /note="assembly_fragment:00298
  fragment_chain:1"
misc_feature
  13170..42396
  /note="assembly_fragment:00928
  fragment_chain:1
  clone_end:SP6
  vector_side:left"
misc_feature
  42497..113471
  /note="assembly_fragment:01958
  fragment_chain:1"
misc_feature
  113572..119459

```

```

/note="assembly_fragment:00442
fragment_chain:2"
misc_feature
  119560..122149
  /note="assembly_fragment:00656
  fragment_chain:2"
misc_feature
  122250..125172
  /note="assembly_fragment:01717
  fragment_chain:2"
misc_feature
  125273..129826
  /note="assembly_fragment:00504
  fragment_chain:3"
misc_feature
  129927..133355
  /note="assembly_fragment:02152
  fragment_chain:3"
misc_feature
  133456..136798
  /note="assembly_fragment:00586
  fragment_chain:4"
misc_feature
  136899..140265
  /note="assembly_fragment:00823
  fragment_chain:4"
misc_feature
  140366..142977
  /note="assembly_fragment:00123"
  143078..146062
  /note="assembly_fragment:00837"
  146163..149355
  /note="assembly_fragment:00879"
  149456..154688
  /note="assembly_fragment:01030"
  154789..162417
  /note="assembly_fragment:01278"
  162518..170626
  /note="assembly_fragment:01660"
  170727..175237
  /note="assembly_fragment:01758"
  175338..177559
  /note="assembly_fragment:01777"
  177660..180444
  /note="assembly_fragment:01980"
  180545..185414
  /note="assembly_fragment:02113"
misc_feature
  /note="assembly_fragment:02113"
BASE COUNT 57924 a 34224 c 34805 g 56330 t 2131 others
ORIGIN

```

Query Match 4.7%; Score 22; DB 2; Length 185414;
 Best Local Similarity 100.0%; Pred. No. 0.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaatgggaca 22
 |||

Db 175922 GAAAGAGAAAGAAATGGGACA 175901

```

RESULT 8
AC024505/c AC024505 HTG 12-MAY-2000
LOCUS Homo sapiens chromosome 1 clone RP11-152L7 map 1, WORKING DRAFT
DEFINITION
AC024505
AC024505.3 GI:7770482
VERSION
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 190856)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 1, clone RP11-152L7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190856)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

```



```

misc_feature      40543..48006
                  /note="assembly-fragment"
misc_feature      48107..53271
                  /note="assembly-fragment"
misc_feature      55372..61182
                  /note="assembly-fragment"
misc_feature      61283..68644
                  /note="assembly-fragment"
misc_feature      68745..75906
                  /note="assembly-fragment"
misc_feature      76007..83998
                  /note="assembly-fragment"
misc_feature      84099..93489
                  /note="assembly-fragment"
misc_feature      93590..101953
                  /note="assembly-fragment"
misc_feature      102054..110821
                  /note="assembly-fragment"
misc_feature      110922..121798
                  /note="assembly-fragment"
misc_feature      121899..133647
                  /note="assembly-fragment"
misc_feature      133748..146812
                  /note="assembly-fragment"
misc_feature      146913..160242
                  /note="assembly-fragment"
misc_feature      160343..175399
                  /note="assembly-fragment"
misc_feature      175500..190856
                  /note="assembly-fragment"
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      58822 a 35587 c 34792 g 58847 t 2808 others
ORIGIN

```

```

Query Match      4.7%; Score 22; DB 2; Length 190856;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 gaagagagaagatgggaca 22
        |||||||
Db      21592 GAAGAGAGAGATGGACA 21571

```

```

RESULT      9
LOCUS      AL162294/c
DEFINITION Human DNA sequence from clone RP5-852L11 on chromosome 20 Contains
ACCESSION   AL162294
VERSION     AL162294.8 GI:8218483
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 61147)
AUTHORS     Lawlor,S.
TITLE       Direct Submission
JOURNAL     Submitted (26-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
          On Jun 3, 2000 this sequence version replaced gi:8217681.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information

```

on the WormPeP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP5-852L11 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP4-800C24 is at 61048 in this sequence. The true right end of clone RP5-981M18 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-852L11 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <http://www.choi1.org/dacpac/home.htm>

VECTOR: pCYPAC2.

FEATURES

source

```

misc_feature      repeat_region
                  /location=Qualifiers
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="20"
                  /clone="RP5-852L11"
                  /clone_id="RPCI-5"
                  complement(411..814)
                  /note="match: GSS: Em:A0020910"
                  611..972
                  /note="LTR1A1 repeat: matches 1..365 of consensus"
                  1057..1793
                  /note="L2 repeat: matches 117..959 of consensus"
                  2150..2341
                  /note="L2 repeat: matches 960..1150 of consensus"
                  2500..2807
                  /note="L2 repeat: matches 1152..1474 of consensus"
                  2865..3227
                  /note="L2 repeat: matches 1563..1966 of consensus"
                  3569..4160
                  /note="MER97a repeat: matches 2..605 of consensus"
                  3752..4381
                  /note="match: GSS: Em:A0055177"
                  4378..4689
                  /note="L1MB7 repeat: matches 5823..6173 of consensus"
                  5457..5504
                  /note="24 copies 2 mer to 95% conserved"
                  5568..5867
                  /note="Alusq repeat: matches 1..300 of consensus"
                  5893..6432
                  /note="MLT2A repeat: matches 1..445 of consensus"
                  6528..6816
                  /note="Alusx repeat: matches 3..299 of consensus"
                  8650..9161
                  /note="MLT2D repeat: matches 58..553 of consensus"
                  9184..9625
                  /note="LTR39 repeat: matches 301..745 of consensus"
                  9626..10095
                  /note="L1PA5 repeat: matches 5672..6141 of consensus"
                  10096..10396
                  /note="LTR39 repeat: matches 2..302 of consensus"
                  11071..11241
                  /note="FRAM repeat: matches 1..170 of consensus"
                  11515..11818
                  /note="Alut repeat: matches 1..310 of consensus"
                  11839..11953
                  /note="L2 repeat: matches 2630..2744 of consensus"
                  12169..12788
                  /note="L1MC4 repeat: matches 7055..7644 of consensus"
                  12789..13090
                  /note="Aluo repeat: matches 1..299 of consensus"

```

```

repeat_region 13091..13705
/note="LIMC4 repeat: matches 6464..7055 of consensus"
repeat_region 13795..13906
/note="L2 repeat: matches 2629..2741 of consensus"
repeat_region 13934..13965
/note="18 copies 2 mer tt 91% conserved"
repeat_region 14474..14969
/note="L2 repeat: matches 2241..2750 of consensus"
repeat_region 16139..16434
/note="AluB repeat: matches 1..289 of consensus"
repeat_region 17275..17405
/note="L2 repeat: matches 2190..2327 of consensus"
repeat_region 17726..18027
/note="AluSg repeat: matches 1..309 of consensus"
repeat_region 18576..18669
/note="LIMB repeat: matches 111..213 of consensus"
repeat_region 18670..19124
/note="LNR2 repeat: matches 3..449 of consensus"
repeat_region 19266..19424
/note="L2 repeat: matches 2344..2504 of consensus"
repeat_region 19479..19559
/note="L2 repeat: matches 2626..2720 of consensus"
repeat_region 19712..19814
/note="MERB1 repeat: matches 1..111 of consensus"
repeat_region 20605..20696
/note="46 copies 2 mer at 66% conserved"
repeat_region 22494..22795
/note="AluB repeat: matches 13..303 of consensus"
repeat_region 23166..23241
/note="MST1 repeat: matches 1..77 of consensus"
repeat_region 23242..23549
/note="AluY repeat: matches 1..308 of consensus"
repeat_region 23550..23874
/note="MST2 repeat: matches 77..394 of consensus"
repeat_region 23945..24183
/note="MIR repeat: matches 11..252 of consensus"
repeat_region 24726..25020
/note="AluSg repeat: matches 1..295 of consensus"
repeat_region 25446..25533
/note="L2 repeat: matches 490..582 of consensus"
repeat_region 25656..25691
/note="MUT1 repeat: matches 112..147 of consensus"
repeat_region 25989..26022
/note="17 copies 2 mer tt 91% conserved"
repeat_region 26209..26874
/note="L2 repeat: matches 803..1526 of consensus"
repeat_region 27517..27802
/note="AluX repeat: matches 1..286 of consensus"
repeat_region 27804..27861
/note="29 copies 2 mer aa 69% conserved"
repeat_region 27888..28082
/note="LIMB3 repeat: matches 5683..5863 of consensus"
misc_feature complement(27922..28443)
/note="match: GSS: Em:AQ431424"
repeat_region 28435..28598
/note="HERV1 repeat: matches 5609..5761 of consensus"
repeat_region 28602..28916
/note="MUT2 repeat: matches 9..328 of consensus"
repeat_region 29097..29394
/note="AluX repeat: matches 1..299 of consensus"
repeat_region 29795..29851
/note="LIMB3 repeat: matches 5998..6054 of consensus"
repeat_region 30088..30360
/note="L2 repeat: matches 2208..2481 of consensus"
repeat_region 33156..34296
/note="LIMC4 repeat: matches 6755..7928 of consensus"
repeat_region 34545..34709
/note="AluY repeat: matches 142..306 of consensus"
repeat_region 34977..35252
/note="AluX repeat: matches 44..312 of consensus"
misc_feature 35650..36023
/note="match: GSS: Em:AQ116392"
repeat_region 37255..37554

```

```

repeat_region 38584..38858
/note="AluX repeat: matches 3..302 of consensus"
repeat_region 39612..39653
/note="THEB repeat: matches 85..364 of consensus"
repeat_region 41265..41480
/note="21 copies 2 mer tt 81% conserved"
repeat_region 41970..42144
/note="MIR repeat: matches 2..236 of consensus"
repeat_region 42789..42907
/note="MERB5 repeat: matches 1..189 of consensus"
repeat_region 43218..43340
/note="L2 repeat: matches 2630..2750 of consensus"
repeat_region 43341..43486
/note="L2 repeat: matches 2623..2750 of consensus"
repeat_region 43677..44054
/note="LIMB8 repeat: matches 6130..6284 of consensus"
repeat_region 44556..46688
/note="Alu repeat: matches 1..302 of consensus"
repeat_region 47390..47990
/note="MUT1 repeat: matches 1..568 of consensus"
repeat_region 48024..48331
/note="AluX repeat: matches 1..304 of consensus"
repeat_region 48343..48561
/note="MIR repeat: matches 30..248 of consensus"
misc_feature 48560..48676
/note="match: GSS: Em:AQ513873"
misc_feature complement(48575..48904)

Query Match 4.5%; Score 21; DB 9; Length 61147;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 atgtgactgctgataaatt 45
Db 14001 Atgtgactgctgataaatt 13981

RESULT 10
AC026323 LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-129P2. WORKING DRAFT SEQUENCE, 15 unordered pieces.
ACCESSION AC026323.9 GI:11094491
VERSION AC026323.9
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 15190)
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbila,J.,
Benton,J., Blinaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowles,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buha,J.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotte,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frenzel,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyet,L.S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Lounsged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

```


TITLE
JOURNAL
AUTHORS
JOURNAL
COMMENT

Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Hawinney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newell, J., Newson, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokweto, S.,
Guh, M., Okunolu, G., Oragun, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojupokan, I., Rolfe, M.,
Ruliz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshari, N.,
Slisdon, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmami, K., Vasquez, L., Vera, V., Vellalton, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, A., Washington, C.,
Washington, S., Williams, G., Williamson, A., Wiczyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
and Gibbs, R.

Unpublished
2 (bases 1 to 155190)

Direct Submission

Submitted (22-MAR-2000)

Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Nov 4, 2000 this sequence version replaced gi:8699803.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: HAPV

Center clone name: RP11-129P2

Summary Statistics

Sequencing vector: MJ3; L08821

Chemistry: Dye-Primer Big Dye; 12% of reads

Chemistry: Dye-terminator Big Dye; 88% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 136343 bases at least Q40

Consensus quality: 141554 bases at least Q30

Consensus quality: 144310 bases at least Q20

Estimated insert size: 144883; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 4.3x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 16 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 43274: contig of 43274 bp in length

43275 43274: gap of unknown length

43275 43275: contig of 25135 bp in length

68510 68509: gap of unknown length

68510 68510: contig of 10272 bp in length

78881 78881: gap of unknown length

78882 78881: contig of 11734 bp in length

90715 90715: gap of unknown length

90716 90715: contig of 8889 bp in length

90816 90816: gap of unknown length

99705 99705: contig of 9553 bp in length

99805 109457: gap of unknown length

109458 118740: contig of 9283 bp in length

118741 118741: gap of unknown length

118841 126400: contig of 7560 bp in length

126401 126500: gap of unknown length

126501 134994: contig of 8494 bp in length

BASE COUNT 47836 a 29970 c 29096 g 46758 t 1530 others

ORIGIN

Query Match 4.5%; Score 21; DB 2; Length 155190;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 atgtgactgcctatgaagt 45
|||||

Db 18469 ATGTGACTGCCTATGAAGT 18489

RESULT 11

LOCUS AC005325

DEFINITION Homo sapiens chromosome 17, clone hRPK.60_A.24, complete sequence.

AC005325

AC005325.1 GI:3366581

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

134995 135094: gap of unknown length
135095 135095: contig of 4463 bp in length
135095 135095: gap of unknown length
139558 139557: gap of unknown length
139558 139557: contig of 3725 bp in length
143881 143882: gap of unknown length
143881 143882: contig of 3964 bp in length
147447 147446: gap of unknown length
147447 147446: contig of 2766 bp in length
150313 150312: gap of unknown length
150313 150312: contig of 2425 bp in length
152838 152837: gap of unknown length
152838 152837: contig of 1098 bp in length
154036 154035: gap of unknown length
154136 154135: contig of 1055 bp in length.

Location/Qualifiers

1. 155190

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="3"

/clone="RP11-129P2"

AC005325 165228 bp DNA PRI 31-JUL-1998
Homo sapiens chromosome 17, clone hRPK.60_A.24, complete sequence.

AC005325

AC005325.1 GI:3366581

HTG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 165228)

1 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

2 (bases 1 to 165228)

TITLE
JOURNAL
COMMENT

Submitted (31-JUL-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 31, 1998 this sequence version replaced gl:3355498.
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

FEATURES

SOURCE

Location/Qualifiers

1. 165228

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_id="RPCK 60_A_24"

/map="17"

/chromosome="17"

/complement(351..650)

/rpt_family="AluSg"

/complement(986..1310)

/rpt_family="MLT1A1"

/complement(1636..1775)

/rpt_family="AluJb"

/complement(1770..2048)

/rpt_family="AluSg"

/complement(2055..2329)

/rpt_family="L1ME"

/complement(2597..2813)

/rpt_family="MERS8A"

/complement(2936..3055)

/rpt_family="FLAM_C"

/complement(3366..3610)

/rpt_family="MIR"

3624..3676

/rpt_family="MLT1J"

3760..3989

/rpt_family="MLT1J"

/complement(5975..6087)

/rpt_family="MLT1J"

8776..9298

/rpt_family="MLT1H"

/complement(9361..9675)

/rpt_family="AluSg"

/complement(9747..10117)

/rpt_family="L1MB6"

10268..10431

/rpt_family="MIR"

11205..11253

/rpt_family="CATn"

13213..13294

/rpt_family="MIR"

13789..13832

/rpt_family="(CA)n"

13833..13867

/rpt_family="polypurine"

/complement(14171..14279)

/rpt_family="MIR"

repeat_region

16367..16650
/rpt_family="AluY"

repeat_region

16651..16688
/rpt_family="(CAAA)n"

repeat_region

16689..16731
/rpt_family="MIR"

repeat_region

16924..16950
/rpt_family="(CAGA)n"

repeat_region

16951..17159
/rpt_family="AluSg"

repeat_region

17166..17436
/complement(17166..17436)

repeat_region

19483
/rpt_family="AluSg"

repeat_region

19484
/rpt_family="MIR"

repeat_region

19785
/rpt_family="AluSx"

repeat_region

19840
/rpt_family="MIR"

repeat_region

20472..20642
/complement(20472..20642)

repeat_region

21141..21142
/rpt_family="MIR"

repeat_region

21143..21144
/rpt_family="AluSx"

repeat_region

21712..21754
/complement(21712..21754)

repeat_region

21755..21946
/rpt_family="Alu"

repeat_region

21947..22395
/complement(21947..22395)

repeat_region

22396..22397
/rpt_family="MIR"

repeat_region

22398..22399
/rpt_family="MIR"

repeat_region

22400..22841
/complement(22400..22841)

repeat_region

22842..24657
/rpt_family="AluSg"

repeat_region

24658..25778
/rpt_family="AluSg"

repeat_region

25779..25805
/rpt_family="AluSx"

repeat_region

25806..26256
/complement(25806..26256)

repeat_region

26257..27073
/rpt_family="AluSg"

repeat_region

27074..27143
/complement(27074..27143)

repeat_region

27144..27306
/rpt_family="MERSA"

repeat_region

27307..27836
/complement(27307..27836)

repeat_region

27837..27967
/rpt_family="MLT1J"

repeat_region

27968..27982
/rpt_family="MER81"

repeat_region

27983..28097
/rpt_family="MER81"

repeat_region

28098..28441
/rpt_family="MER5A"

repeat_region

28442..28556
/rpt_family="L2"

repeat_region

28557..30757
/rpt_family="L2"

repeat_region

30758..31192
/rpt_family="(TGAA)n"

repeat_region

31193..31351
/rpt_family="L2"

repeat_region

31352..31434
/rpt_family="(TGCA)n"

repeat_region

31435..31615
/complement(31435..31615)

repeat_region

31616..31957
/rpt_family="A-rich"

repeat_region

31958..32091
/complement(31958..32091)

repeat_region

32092..32153
/rpt_family="AluY"

repeat_region

32154..33172
/rpt_family="MLT1J"

repeat_region

33173..33226
/rpt_family="purine-rich"

repeat_region

33227..33303
/complement(33227..33303)

repeat_region

33304..33610
/rpt_family="L2"

repeat_region

33611..34078
/complement(33611..34078)

repeat_region

34079..34245
/complement(34079..34245)

RESULT 14
HSDJ130E4 108893 bp DNA RPL 13-APR-2000
LOCUS Human DNA sequence from clone RPL-130E4 on chromosome 6q24.2-25.3.
DEFINITION Contains the 3' end of the ESR1 gene for estrogen receptor 1, the
3' end of the gene KIAA0796, ESTs, STS and GSSs, complete
sequence.
ACCESSION AL078582 GI:7161747
VERSION AL078582.13
KEYWORDS HTG: ESR1; estrogen receptor; KIAA0796.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 108893)
AUTHORS Parker, A.
TITLE Direct Submission
JOURNAL Submitted (06-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clone@sanger.ac.uk
On Mar 6, 2000 this sequence version replaced gi:7018350.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/_C_elegans/wormpep This sequence
was generated from part of bacterial clone configs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RPL-130E4 is from the library RPL-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone
RPL-130E4. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RPL-130E4 is at 108893 in this
sequence. The true right end of clone RPL-6515 is at 100 in this
sequence.

FEATURES
source
1..108893
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q24.2-25.3"
/clone="RPL-130E4"
/clone_lib="RPL-1"
249..932
misc_feature
/note="match: GSS: Em:AQ373843"
1912..1967
repeat_region
/note="2 copies 28 mer 98% conserved"
2415..2462
repeat_region
/note="8 copies 6 mer caaaa 81% conserved"
2418..2463
repeat_region
/note="23 copies 2 mer aa 80% conserved"
3009..3407
/note="MSB repeat: matches 1..426 of consensus"
3449..3681
repeat_region
/note="MER58A repeat: matches 1..214 of consensus"

repeat_region
4248..4299
/note="13 copies 4 mer ctcc 73% conserved"
repeat_region
4453..4742
/note="5 copies 58 mer 76% conserved"
repeat_region
4460..4733
/note="137 copies 2 mer tc 77% conserved"
repeat_region
4461..4732
/note="68 copies 4 mer ctct 77% conserved"
repeat_region
4463..4732
/note="45 copies 6 mer ctctct 78% conserved"
repeat_region
4467..4746
/note="10 copies 28 mer 76% conserved"
repeat_region
4476..4731
/note="2 copies 128 mer 82% conserved"
repeat_region
6026..6404
/note="LIM6 repeat: matches 5497..5874 of consensus"
7053..7148
/note="LIM6 repeat: matches 29..124 of consensus"
repeat_region
8742..8789
/note="MIR repeat: matches 29..124 of consensus"
repeat_region
9564..9615
/note="24 copies 2 mer ac 91% conserved"
repeat_region
9743..9866
/note="MIR repeat: matches 186..237 of consensus"
repeat_region
10164..10258
/note="L2 repeat: matches 2576..2709 of consensus"
repeat_region
10427..10669
/note="LIM8 repeat: matches 6076..6171 of consensus"
misc_feature
11871..12398
/note="LIM4 repeat: matches 5115..5369 of consensus"
repeat_region
11997..12046
/note="match: GSS: Em:AQ439940"
repeat_region
/note="25 copies 2 mer tt 74% conserved"
repeat_region
/note="ESR1"
/gene="ESR1"
/note="match: cDNAs: Em:AF124093 Em:AB007453 Em:M12674
Em:X03635 Em:AF042060 Em:Y00102 Em:AF042061 Em:U75604
Em:AF061275 Em:U75605 Em:AJ242740 Em:AJ242741 Em:AF119229
Em:AF095911 Em:X61098 Em:AF110402 Em:AB003356 Em:AB012721
Em:X38651 Em:Y18017 Em:AF099079 Em:AF136980 Em:Z37167
Em:AF172069 Em:AF172068 Em:AB036415 Em:AF182220
Em:AF182221 Em:AF172068 Em:AF177465 Em:AF113513
Em:AF181077 Em:Z75126 Em:Z49257 Em:U47678 Em:X31559
Em:AF177936 Em:AF136979 Em:D28954 Em:X89959
match: ESTs: Em:AL040974 Em:A1634735 Em:A1584840
Em:A1888116 Em:A1675862 Em:AM427925 Em:A1978964
Em:AA336802 Em:AW169860 Em:A1810941 Em:A1654044
Em:A1202659 Em:AA296578 Em:AA337245 Em:AA420328
Em:A1952766 Em:AV175085 Em:AM427924 Em:AM089672
Em:A1972926 Em:A1669577 Em:A1983175 Em:AA291702
Em:AM088459 Em:AM438839 Em:AA319335 Em:AA164585
Em:A1654018 Em:AA164586 Em:A1625626 Em:T28397 Em:A1673252
Em:AA336938 Em:AA336971 Em:A1540656 Em:AA299347
Em:A1862012 Em:AL040975 Em:AN018232 Em:AM440601
Em:A1669190 Em:AA337706 Em:AA331020 Em:A1888281 Em:C02879
Em:A1984079"
/evidence="not_experimental"
/product="dJ130E4.1 (estrogen receptor 1)"
12052..54335
/gene="ESR1"
join(<12052..12185,45446..45629,49793..50027)
/gene="ESR1"
/note="match: proteins: SW:P03372 Tr:Q95131 SW:P81559
Sw:P06211 Sw:P06212 Sw:P19785 Sw:Q91250 Tr:Q9YHM3
Sw:Q29040"
/evidence="not_experimental"
/product="dJ130E4.1 (estrogen receptor 1)"
/protein_id="CA87585.1"
/db_xref="GI:7573520"
/translation="NOGKCEGMEVLEFDMLATSSRPRMNLGEEFVCIKSIILNS
GYVTFSLTKSLSEEDHTRVLDITDTLHLMKAGCTQQOQORLAQLITISHI
RMSNKGMEHLSMKCKNVPLYDLLEMLDHRHAPTSRGAASVEETDOSHATAG
SSSHSLQKTYITGAEKFPATV"

```

repeat_region 12791..12834 /note="22 copies 2 mer lt 81% conserved"
repeat_region 12796..12831 /note="6 copies 6 mer titltg 88% conserved"
repeat_region 13192..13447 /note="2 copies 128 mer 92% conserved"
repeat_region 14237..14286 /note="2 copies 25 mer 96% conserved"
repeat_region 15390..15551 /note="L2 repeat: matches 2364..2537 of consensus"
repeat_region 15599..15658 /note="L2 repeat: matches 2702..2750 of consensus"
misc_feature complement(17729..18056) /note="match: GSS: Em:AQ061898"
repeat_region 17821..18050 /note="MIR repeat: matches 23..252 of consensus"
misc_feature 17877..18256 /gene="ESR1"
repeat_region 18641..18916 /note="match: GSS: Em:AQ052854"
repeat_region 19498..19640 /note="AluSq repeat: matches 37..306 of consensus"
repeat_region 20105..20541 /note="MER5B repeat: matches 1..140 of consensus"
repeat_region 20573..20927 /note="LFR26 repeat: matches 151..601 of consensus"
repeat_region 22022..22071 /note="MER1B repeat: matches 1..337 of consensus"
repeat_region 22029..22082 /note="MIR repeat: matches 208..257 of consensus"
repeat_region 22087..22152 /note="L2 repeat: matches 2639..2692 of consensus"
repeat_region 22449..22639 /note="MIR repeat: matches 116..182 of consensus"
repeat_region 24767..24850 /note="MIR repeat: matches 8..224 of consensus"
repeat_region 25014..26152 /note="MIR repeat: matches 22..100 of consensus"
misc_feature 26439..26977 /note="L1MK1 repeat: matches 5162..6324 of consensus"
misc_feature 26664..27171 /note="match: GSS: Em:AQ67837"
misc_feature 26665..27078 /note="match: GSS: Em:AQ081793"
misc_feature 27152..27515 /note="match: GSS: Em:AQ825841"
repeat_region 27516..29089 /note="THE1A repeat: matches 1..354 of consensus"
repeat_region 29090..29326 /note="THE1A-internal repeat: matches 1..1580 of consensus"
repeat_region 29566..34858 /note="THE1A repeat: matches 1..238 of consensus"

Query Match 4.3%; Score 20; DB 9; Length 108893;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaatgga 20
|||||
Db 4581 GAAGAGAAAGAGATGGGA 4562

RESULT 15
PFMAL4P1_0
WPCOMMENT
Sequence split into 4 fragments LOCUS PFMAL4P1 Accession AL034557
Fragment Name Begin End
PFMAL4P1_0 1 110000

```

```

PFMAL4P1_1 100001 210000
PFMAL4P1_2 200001 310000
PFMAL4P1_3 300001 392633
LOCUS PFMAL4P1 392633 bp DNA HTG 11-AUG-1999
DEFINITION Plasmodium falciparum chromosome 4 strain 3D7, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL034557.7 GI:5731897
VERSION HTG; HTGS_PHASE1.
KEYWORDS malaria parasite P. falciparum.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 392633)
AUTHORS Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
and Barrell,B.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) P. falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
COMMENT On Aug 12, 1999 this sequence version replaced gi:5531346.
For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
sequence is unfinished and does not necessarily represent the
correct sequence. Work on the sequence is in progress and the
release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
Order of segments is not known; 800 n's separate segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source 1..392633
location/Qualifiers
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="4"
BASE COUNT 146453 a 41490 c 42490 g 138992 t 23208 others
ORIGIN
Query Match 4.3%; Score 20; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaatgga 20
|||||
Db 30271 GAAGAGAAAGAGATGGGA 30290

```

Search completed: April 6, 2002, 06:18:43
 Job time: 21950 sec

Sat Apr 6 11:03:26 2002

us-09-621-781-1_copy_1507_1970.rge

Page 14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2002, 04:29:53 ; Search time 157.36 Seconds
(without alignments)
2527.954 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Sequence: 1 gaaagagaagaagaatggya.....tttgcaaacggaaccgac 464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

word size : 0

Total number of hits satisfying chosen parameters: 1861242

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database :

1:	/SID52/gcgdata/geneseq/geneseqn/NA1980.DAT *
2:	/SID52/gcgdata/geneseq/geneseqn/NA1981.DAT *
3:	/SID52/gcgdata/geneseq/geneseqn/NA1983.DAT *
4:	/SID52/gcgdata/geneseq/geneseqn/NA1984.DAT *
5:	/SID52/gcgdata/geneseq/geneseqn/NA1985.DAT *
6:	/SID52/gcgdata/geneseq/geneseqn/NA1986.DAT *
7:	/SID52/gcgdata/geneseq/geneseqn/NA1987.DAT *
8:	/SID52/gcgdata/geneseq/geneseqn/NA1988.DAT *
9:	/SID52/gcgdata/geneseq/geneseqn/NA1989.DAT *
10:	/SID52/gcgdata/geneseq/geneseqn/NA1990.DAT *
11:	/SID52/gcgdata/geneseq/geneseqn/NA1991.DAT *
12:	/SID52/gcgdata/geneseq/geneseqn/NA1992.DAT *
13:	/SID52/gcgdata/geneseq/geneseqn/NA1993.DAT *
14:	/SID52/gcgdata/geneseq/geneseqn/NA1994.DAT *
15:	/SID52/gcgdata/geneseq/geneseqn/NA1995.DAT *
16:	/SID52/gcgdata/geneseq/geneseqn/NA1996.DAT *
17:	/SID52/gcgdata/geneseq/geneseqn/NA1997.DAT *
18:	/SID52/gcgdata/geneseq/geneseqn/NA1998.DAT *
19:	/SID52/gcgdata/geneseq/geneseqn/NA1999.DAT *
20:	/SID52/gcgdata/geneseq/geneseqn/NA2000.DAT *
21:	/SID52/gcgdata/geneseq/geneseqn/NA2001.DAT *
22:	/SID52/gcgdata/geneseq/geneseqn/NA2001.DAT *

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	223	48.1	2614	19	AAV65766	Rat progression el
2	223	48.1	2614	20	AA23030	Rat progression el
3	108	23.3	2111	19	AAV65765	Human progression
4	108	23.3	2111	20	AA23029	Human progression
5	108	23.3	2111	20	AA878383	Human progression
6	199	4.1	72928	20	AA18355	Human ASTHJ 5' ge
7	19	4.1	72928	21	AA80253	Human ASTHJ 5' ge
8	18	3.9	819	22	AAH05417	Human cDNA clone (
9	17	3.7	686	20	AAV14334	H. pylori GHPD 212
10	17	3.7	1288	21	AAV26436	Human secreted pro
11	17	3.7	1278	21	AAV76612	Human ORFX ORF2167

12	17	3.7	1648	22	AAC87123	Nucleotide sequenc
13	17	1775	22	AAC87122	Nucleotide sequenc	
14	17	3.7	2075	21	AAC44199	Aradiopsis thalia
14	17	3.7	3122	16	AAO79913	Lys-aminopeptidase
15	17	3.7	3591	22	AAH16646	Human CDNA sequenc
16	17	3.7	3729	22	AAE77688	Human wild-type Fc
17	17	3.7	4131	22	AAE77689	Human variant Fc
18	17	3.7	5538	22	AAH18660	Human CDNA sequenc
19	17	3.7	11298	18	AAE68756	Human high affinity
20	17	3.7	11298	19	AAV54661	Human beta subunit
21	17	3.7	11298	21	AAE20937	Human high affinity
22	17	3.7	11298	21	AAH38115	Human adenosine re
22	17	3.7	11298	21	AAAF9214	Human IGBB gene S
23	17	3.7	11298	22	AAAF9214	Human FcEbeta beta C
24	17	3.7	11298	22	AAAF9214	Human high affinity
25	17	3.7	11357	14	AAO51024	Human adenosine re
26	17	3.7	21742	21	AAE20938	Human SC3 DNA. Ho
27	17	3.7	21742	21	AAH34816	Human Atopy related gene
28	17	3.7	28720	19	AAV49655	Human receptor-rel
29	17	3.7	33030	22	AAE29337	Human secreted pro
30	17	3.7	117609	21	AAE11435	Human secreted pro
31	16	3.4	63	21	AAAC11216	Human secreted pro
32	16	3.4	114	22	AAI146005	Human prostate can
33	16	3.4	161	21	AAAC26461	Expressed Sequence
34	16	3.4	232	14	AAO38840	Human Brain Expres
35	16	3.4	253	14	AAO38840	Human prostate can
36	16	3.4	253	21	AAAC26462	Human prostate can
37	16	3.4	287	21	AAAC09789	Human brain specifi
38	16	3.4	296	22	AAH57333	Human brain specifi
39	16	3.4	340	22	AAI12552	Probe #12485 for g
40	16	3.4	340	22	AAI147842	Probe #16528 used
41	16	3.4	340	22	AAI10823	Probe #8226 used t
42	16	3.4	341	22	AAE91993	B thuringiensis 14
43	16	3.4	389	22	AAH57541	Human brain cell s
44	16	3.4	417	20	AAV88497	EST clone GT6. Ho
45	16	3.4	416	21	AAAC29007	Human secreted pro

ALIGNMENTS

RESULT	1
AAV65766	
ID	AAV65766 standard; cDNA; 2614 BP.
XX	
AC	AAV65766;
XX	
DT	02-FEB-1999 (first entry)
XX	
DE	Rat progression elevated gene-2 (PEG-3) promoter region.
XX	
KW	Progression elevated gene-3; PEG-3; rat; tumour progression;
KW	DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
KW	brain cancer; cervix cancer; prostate cancer; lung cancer;
KW	colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy;
KW	promoter; ss.
OS	
XX	
OS	Rattus sp.
XX	
FH	
FT	Key
FT	location/Qualifiers
FT	128..133
FT	/*tag= a
FT	/label= PEA3
FT	2128..2133
FT	/*tag= b
FT	/label= PEA3
FT	2404..2409
FT	/*tag= c
FT	/label= PEA3
FT	338..344
FT	/*tag= d
FT	/label= E2A
FT	351..357
FT	/*tag= e
FT	/label= GRE

FT	misc_feature	1094..1099	/tag= f
FT		/label= GRE	
FT	misc_feature	729..737	/tag= g
FT		/label= E2F	
FT	misc_feature	193..197	/tag= h
FT		/label= TRE	
FT	misc_feature	489..493	/tag= i
FT		/label= TRE	
FT	misc_feature	936..940	/tag= j
FT		/label= TRE	
FT	misc_feature	774..779	/tag= k
FT		/note= "acute phase reactive regulating element"	
FT	misc_feature	2090..2095	/tag= l
FT		/note= "acute phase reactive regulating element"	
FT	misc_feature	2028..2033	/tag= m
FT		/label= SPI	
FT	misc_feature	1784..1790	/tag= n
FT		/label= Ap1	
FT	misc_feature	2552..2558	/tag= o
FT		/label= Ap1	
FT	misc_feature	865..874	/tag= p
FT		/label= Ap2	
FT	misc_feature	1597..1609	/tag= q
FT		/label= NF1	
XX	W09842315-A1.		
XX			
PD	01-OCT-1998.		
XX			
PF	20-MAR-1998;	98WO-US05793.	
XX			
PR	21-MAR-1997;	97US-0821818.	
XX			
PA	(GENO-) GENQUEST INC.		
PA	(UTCO) UNIV COLUMBIA NEW YORK.		
XX			
XX	Fisher PB, Goldstein NI, Su Z, Zhang N;		
DR	WPI; 1998-557025/47.		
XX			
PT	New isolated Progression Elevated Gene-3 - used to develop products		
PT	for e.g. modulating DNA damage and repair pathways, cancer		
XX	progression or oncogene mediated transformation and angiogenesis.		
XX	Claim 24; Fig 14A-B; 225pp; English.		

This is the 5' regulatory region of the rat progression elevated gene-3 (PEG-3, see also AAW65764). It was identified using a genomic walking strategy, and appears to encompass a functionally complete PEG-3 gene promoter including a number of potentially important transcriptional motifs. PEG-3 expression is uniquely elevated in all cases of rodent progression analysed, and the level of expression is also an indicator of DNA damage in that cell. Cells in which a reporter gene is under control of the promoter of the PEG-3 gene can be used in methods for identifying agents that modulate PEG-3 expression or the ability of PEG-3 to induce progression, or for determining whether an agent is capable of inhibiting DNA damage and repair pathways, cancer progression or oncogene-mediated transformation. Compounds that induce DNA damage or which regulate angiogenesis can also be identified using such cells. The methods can be applied to a progression phenotype

CC		comprising anchorage-independent growth, tumorigenesis, angiogenesis
CC		or metastasis to melanoma, brain, cervical, prostate, lung or
CC		colorectal cancer, neuroblastoma or glioblastoma.
XX		
SQ	Sequence 2614 BP; 737 A; 615 C; 706 G; 556 T; 0 other;	
	Query Match	48.1%; Score 223; DB 19; Length 2614;
	Best Local Similarity	100.0%; Pred. No. 8.5e-110;
	Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps	
QY	1 gaaagagaagaaggatggaacagcatgtgactgcctgatgaagtggcgtgtccca 60	
Db	1512 gaagaagaagaaggatggaacagcatgtgactgcctgatgaagtggcgtgtccaa 1571	
QY	61 aagttctcgagatttgacgcgtctcttgattgaccagaacacgccttggagaagccag 120	
Db	1572 aagttctcgagatttgacgcgtctcttgattgaccagaacacgccttggagaagccag 1631	
QY	121 gtagcttcacaagcccggaatctccgcggaatttaagtgttgtttctctccacc 180	
Db	1632 gtgacttcacaagcccggaatctccgcggaatttaagtgttgtttctctccacc 1691	
QY	181 ttctcagggaacttcgaaactccgcctctccggtgagcgtcacg 223	
Db	1692 ttctcagggaacttcgaaactccgcctctccggtgagcgtcacg 1734	
RESULT 2		
AAZ23030		
ID	AAZ23030 standard; DNA; 2614 BP.	
XX		
AC	AAZ23030;	
XX		
DT	17-JAN-2000 (first entry)	
DE	Rat progression elevated gene-3 (PGC-3) promoter sequence.	
XX		
KW	Progression-elevated gene-3; PGC-3 gene; regulatory region; cell death;	
KW	cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; rat;	
KW	promoter; ss.	
XX		
OS	Rattus sp.	
XX		
PN	M0994898-A1.	
PD	07-OCT-1999.	
PF	31-MAR-1999; 99MO-US07199.	
PR	31-MAR-1998; 98US-0052753.	
PA	(UYCO) UNIV COLUMBIA NEW YORK.	
P1	Fisher PB:	
PI	WPI; 1999-591184/50.	
DR	Novel vectors useful for studying the progression of cancer -	
PT		
PS	Disclosure; Fig 14 A-B; 251pp; English.	
XX		
XX		
CC	The invention relates to an inducible progression-elevated gene-3 (PGC-3	
CC	gene) regulatory region functionally linked to a gene encoding a product	
CC	that causes or may be induced to cause the death or inhibition of cancer	
CC	cell growth. A vector of the invention which contains a gene encoding	
CC	thymidine kinase or a product which causes the cell to express a	
CC	specific antigen can be administered along with gancyclovir or acyclovir,	
CC	or an antibody or fragment to the antigen, respectively, to treat cancer	
CC	in a subject. The PGC-3 gene is useful for generating new cloning and	
CC	expression vectors, transfected cells, and for developing methods for	
CC	cultured growth of such cells. The PGC-3 polynucleotide is also useful	
CC	as a source of primers and probes to study the progression of cancer, and to	

CC detect the presence of the gene. The present sequence represents the
CC nucleotide sequence of the rat PEG-3 gene promoter.
XX
SQ Sequence 2614 BP; 737 A; 615 C; 706 G; 556 T; 0 other;

Query Match 48.1%; Score 223; DB 20; Length 2614;
Best Local Similarity 100.0%; Pred. No. 8.5e-110;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaagaagaagaatgagacagcatgtgactgacctgaatgaagtgtgctgtctcaaa 60
|||||
Db 1512 gaagaagaagaatgagacagcatgtgactgacctgaatgaagtgtgctgtctcaaa 1571
QY 61 aagttctgagagattgacgctctctgagattgaagccaagaacgctctgggaagccag 120
|||||
Db 1572 aagttctgagagattgacgctctctgagattgaagccaagaacgctctgggaagccag 1631
QY 121 gtgacctcacaaagcccgagatctccgagaaatttcagtggttttctctctccacc 180
|||||
Db 1632 gtgacctcacaaagcccgagatctccgagaaatttcagtggttttctctctccacc 1691
QY 181 ttctcaggagactccgaactccgctctccggtgagctcag 223
|||||
Db 1692 ttctcaggagactccgaactccgctctccggtgagctcag 1734

RESULT- 3
AAV65765 standard; cDNA; 2111 BP.
ID AAV65765 standard; cDNA; 2111 BP.
XX
AC AAV65765;

DT 02-FEB-1999 (first entry)

XX Human progression elevated gene-2 (PEG-3) cDNA.

XX Progression elevated gene-3; PEG-3; human; tumour progression;
KM DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
KM brain cancer; cervix cancer; prostate cancer; lung cancer;
KM colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy;
ss.

OS Homo sapiens.

XX Key Location/Qualifiers
FH 294..2030
FT CDS /*tag= a
FT

XX MO9842315-A1.

XX 01-OCT-1998.

XX 20-MAR-1998; 98WO-US05793.

XX 21-MAR-1997; 97US-0821818.

XX (GENO-) GENQUEST INC.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Fisher PB, Goldstein NI, Su Z, Zhang N;

XX WPI: 1998-557025/47.

XX P-PSDB; AAW79958.

XX New isolated Progression Elevated Gene-3 - used to develop products
PT for e.g. modulating DNA damage and repair pathways, cancer
PT progression or oncogene mediated transformation and angiogenesis.

XX Claim 4; Fig 13A-C; 225pp; English.

XX This nucleotide sequence includes an open reading frame encoding
CC human progression elevated gene-3 (PEG-3) protein (see AAW79957).

CC PEG-3 clones were isolated from an MCF-7 cDNA library using rat
CC PEG-3 cDNA (see AAV65764) as probe. 5' RACE (see AAV65767-68) was used
CC to generate full-length cDNA. PEG-3 mRNA is highly expressed in
CC most human tumour cell lines. Probes derived from the gene can be
CC used to monitor tumour progression. Antisense sequences can be
CC used to inhibit tumour progression. Cells in which a reporter gene
CC is under control of the promoter (see AAV65766) of the rat PEG-3 gene
CC can be used in methods for identifying agents that modulate PEG-3
CC expression or the ability of PEG-3 to induce progression, or for
CC determining whether an agent is capable of inhibiting DNA damage
CC and repair pathways, cancer progression or oncogene-mediated
CC transformation. Compounds that induce DNA damage or which regulate
CC angiogenesis can also be identified using such cells. Transgenic
CC animals and vaccines comprising PEG-3 polypeptides and an immune
CC response enhancer are also claimed. The methods can be applied to
CC a progression phenotype comprising anchorage-independent growth,
CC tumorigenesis, angiogenesis or metastasis, to melanoma, brain,
CC cervical, prostate, lung or colorectal cancer, neuroblastoma or
CC glioblastoma.

SQ Sequence 2111 BP; 533 A; 607 C; 575 G; 396 T; 0 other;

Query Match 23.3%; Score 108; DB 19; Length 2111;
Best Local Similarity 100.0%; Pred. No. 4.2e-48;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 ccttgcgggagacgcttgcagagccggtgagacatcagctccgagcccaagc 391
|||||
Db 120 ccttgcgggagacgcttgcagagccggtgagacatcagctccgagcccaagc 179
QY 392 tgaggcgacatgaacgcgctgacctgagagacaatccggaaccaca 439
|||||
Db 180 tgaggcgacatgaacgcgctgacctgagagacaatccggaaccaca 227

RESULT 4
AAZ23029

ID AAZ23029 standard; cDNA; 2111 BP.

XX AAZ23029;

XX 17-JAN-2000 (first entry)

XX Human progression elevated gene-3 (PEG-3) cDNA sequence.

XX Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;

XX cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; human; ss.

XX Homo sapiens.

XX MO9949898-A1.

XX 07-OCT-1999.

XX 31-MAR-1999; 99WO-US07199.

XX 31-MAR-1998; 98US-0052753.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Fisher PB;

XX WPI: 1999-591184/50.

XX P-PSDB; AAV41104.

XX Novel vectors useful for studying the progression of cancer -
PT Disclosure; Fig 13A-C; 251pp; English.

XX The invention relates to an inducible progression-elevated gene-3 (PEG-3
CC gene) regulatory region functionally linked to a gene encoding a product
CC that causes or may be induced to cause the death or inhibition of cancer

CC cell growth. A vector of the invention which contains a gene encoding
CC thymidine kinase or a product which causes the cell to express a
CC specific antigen can be administered along with gancyclovir or acyclovir,
CC or an antibody or fragment to the antigen, respectively, to treat cancer
CC in a subject. The PEG-3 gene is useful for generating new cloning and
CC expression vectors, transfected cells, and for developing methods for
CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as
CC a source of primers and probes to study the progression of cancer, and to
CC detect the presence of the gene. The present sequence represents the
CC nucleotide sequence of the human PEG-3 cDNA.
XX
SQ Sequence 2111 BP; 533 A; 608 C; 575 G; 395 T; 0 other;

Query Match 23.3%; Score 108; DB 20; Length 2111;
Best Local Similarity 100.0%; Pred. No. 4.2e-48;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 332 cctttgcgggacagcctttgagacagccgtgagacatcacgctcccgagcccaagcc 391
b 120 cctttgcgggacagcctttgagacagccgtgagacatcacgctcccgagcccaagcc 179
QY 392 tgaaggcgacatgaacgcgctgagacagcattgagacatccgagcccaagca 439
Db 180 tgaaggcgacatgaacgcgctgagacagcattgagacatccgagcccaagca 227

RESULT 5
AAZ87383
ID AAZ87383 standard; cDNA; 2111 BP.
XX
AC AAZ87383;
XX
DT 08-OCT-1999 (first entry)
XX
DE Human progression elevated gene-3 (PEG-3) cDNA.
XX
KM Progression elevated gene-3; PEG-3; human; angiogenesis; cancer;
XX therapy; vaccine; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 294..2030
FT CDS /*tag- a
XX
PN WO937776-A1.
XX
PD 29-JUL-1999.
XX
PF 26-JAN-1999; 99WO-US01623.
XX
PR 29-JAN-1998; 98US-0072941.
PR 26-JAN-1998; 98US-0072469.
XX
PA (GENO-) GENQUEST INC.
XX
PI Fisher PB, Zhang N;
XX
DR WPI: 1999-458694/38.
DR P-PSDB; AAY06514.
XX
XX

Modulation of angiogenesis by altering the expression and/or
activity of a progression-associated protein, especially for cancer
treatment

Claim 1; Page 71-74; 81pp; English.

This is the nucleotide sequence of a cDNA clone corresponding to
the human progression elevated gene-3 (PEG-3). It codes for a
578-amino acid protein (see AAY06514), and represents a new member
of the gadd34/MyD16 gene family. The human PEG-3 cDNA was
isolated from a MCF-7 cDNA library using rat PEG-3 (see AAZ87387) as

CC probe. Human PEG-3 is generally expressed in cells that are in
CC progression, including most human tumour cell lines. A claimed
CC method for modulating angiogenesis in an organism comprises
CC administering an agent that alters expression and/or activity of a
CC PEG-3 protein, such as an antisense polynucleotide or antibody.
CC Also claimed is a method for determining whether an agent
CC modulates angiogenesis, where the candidate agent is present
CC within a combinatorial small molecule library. Vaccines and
CC pharmaceutical compositions comprising such compounds are also
CC provided and may be used to prevent angiogenesis, especially
CC related to cancer cell progression.
XX
SQ Sequence 2111 BP; 533 A; 608 C; 575 G; 395 T; 0 other;

Query Match 23.3%; Score 108; DB 20; Length 2111;
Best Local Similarity 100.0%; Pred. No. 4.2e-48;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 332 cctttgcgggacagcctttgagacagccgtgagacatcacgctcccgagcccaagcc 391
Db 120 cctttgcgggacagcctttgagacagccgtgagacatcacgctcccgagcccaagcc 179
QY 392 tgaaggcgacatgaacgcgctgagacagcattgagacatccgagcccaagca 439
Db 180 tgaaggcgacatgaacgcgctgagacagcattgagacatccgagcccaagca 227

RESULT 6
AAZ18355
ID AAZ18355 standard; DNA; 72928 BP.
XX
AC AAZ18355;
XX
DT 19-OCT-1999 (first entry)
XX
DE Human ASTH1J 5' genomic region.
XX
KM ASTH1; asthma; human; chromosome 11p; ASTH1J; ASTH1J; genetic locus;
XX therapeutic; immunogen; ds.
OS Homo sapiens.
XX
FH WO937809-A1.
FT CDS 294..2030
FT CDS /*tag- a
XX
PN 29-JUL-1999.
XX
PF 21-JAN-1998; 98WO-US01260.
XX
PR 21-JAN-1998; 98WO-US01260.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Brooks-Wilson AR, Buckler A, Cardon L, Carey AH;
PI Galvin M, Miller A, North M;
XX
DR WPI: 1999-479058/40.
XX
XX
PT Mammalian asthma related genes, useful for diagnosis of a
PT predisposition to development of asthma
XX
PS Claim 15; Page 75-96; 195pp; English.

The invention identifies a genetic locus ASTH1, associated with asthma,
mapped to human chromosome 11p. ASTH1 and ASTH1J are genes present
within the locus, located close to each other on human chromosome 11p,
and have similar patterns of expression, and common sequence motifs. The
ASTH1 genes and fragments, encoded protein, genomic regulatory regions
and anti-ASTH1 antibodies are useful in the identification of individuals
CC predisposed to development of asthma, and for the modulation of gene
CC activity in vivo for prophylactic and therapeutic purposes. The ASTH1
CC protein is useful as an immunogen to raise specific antibodies, in drug
screening for compositions that mimic or modulate ASTH1 activity or

CC expression, including altered forms of ASTH1 protein, and as a
CC therapeutic. The present sequence represents a human ASTH1 genomic region
CC sequence.

CC Sequence 72928 BP; 20042 A; 15100 C; 16036 G; 21750 T; 0 other;

XX SQ

Query Match

Best Local Similarity 4.1%; Score 19; DB 20; Length 72928;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 aaagagaagaagatggga 20

DB 46222 aaagagaagaagatggga 46240

RESULT 7

AAA80253

AAA80253 standard; DNA; 72928 BP.

XX AC

AAA80253;

XX DT

22-NOV-2000 (first entry)

XX DE

Human ASTH1J 5' genomic region.

XX KW

ASTH1 locus; ASTH1J; human; chromosome 11p; asthma;

XX KW

bronchial hyperreactivity; ets family; transcription factor;

XX KW

splice variant; genetic predisposition; polymorphism; antibody;

XX KW

drug screening; prophylaxis; therapy; diagnosis; ds.

XX OS

Homo sapiens.

XX AC

US6087485-A.

XX PD

11-JUL-2000.

XX PF

21-JAN-1998; 98US-0009913.

XX PR

21-JAN-1997; 97US-0035663.

XX PR

01-JUL-1997; 97US-0051432.

XX PA

(AXYS-) AXYS PHARM INC.

XX PI

Galvin M. Miller A. North M. Cardon L. Buckler A.

XX PI

Brooks-Wilson AR, Carey AH;

XX DR

WPI; 2000-505109/45.

XX PT

New nucleic acids other than naturally occurring chromosomes encoding

XX PT

ASTH1 protein, for e.g. screening compositions that modulate expression

XX PT

or function of ASTH1 proteins or as diagnostics for genetic

XX PT

predisposition to asthma

XX PS

Claim 7; Column 49-112; 131pp; English.

XX CC

The invention relates to the ASTH1 locus on the short arm of human

XX CC

chromosome (11p). This locus comprises the ASTH1 and ASTH1J genes,

XX CC

which are associated with a genetic predisposition to asthma and

XX CC

bronchial hyperreactivity. The ASTH1 and ASTH1J genes are oriented in

XX CC

opposite directions with the ASTH1 locus, and have similar patterns of

XX CC

expression and common sequence motifs. They are both expressed in

XX CC

trachea, lung and several other tissues. ASTH1 and ASTH1J are novel

XX CC

members of the ets family of transcription factors, which have been

XX CC

implicated in the activation of a variety of genes including the TCRA

XX CC

gene and cytokine genes known to be important in the aetiology of asthma.

XX CC

Both ASTH1 and ASTH1J mRNAs are alternatively spliced. Alternative

XX CC

splicing of transcripts has no effect on the open reading frame of

XX CC

ASTH1J, as the exons involved are all 5' to the start codon in exon b. In

XX CC

contrast, alternative splicing of ASTH1 transcripts results in 3

XX CC

different ASTH1 isoforms. The invention also encompasses mouse asth1j

XX CC

protein. The ASTH1 nucleic acids are useful as diagnostics to identify a

XX CC

hereditary predisposition to asthma, as probes for identifying ASTH1

CC related genes, for identifying expression of the gene in a biological

CC specimen, and for generating genetically modified non-human animals or

CC site specific gene modifications in cell lines. The encoded ASTH1

CC proteins are useful as immunogens to raise specific antibodies; in drug

CC screening for compositions that mimic or modulate activity or expression

CC of ASTH1 and/or ASTH1J (including altered forms of these proteins); and

CC as a therapeutic. The ASTH1 genes or fragments thereof, encoded proteins,

CC ASTH1 genomic regulatory regions, and anti-ASTH1 and anti-ASTH1J

CC antibodies are useful in the identification of individuals predisposed to

CC development of asthma, and for modulation of gene activity in vivo for

CC prophylactic and therapeutic purposes. The intact ASTH1 or ASTH1J

CC proteins or active fragments thereof may be used to modulate or reduce

CC bronchial hyperreactivity. The present sequence represents the

CC non-transcribed region upstream of the human ASTH1J gene.

XX SQ

Sequence 72928 BP; 20041 A; 15101 C; 16036 G; 21750 T; 0 other;

Query Match

Best Local Similarity 4.1%; Score 19; DB 21; Length 72928;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 aaagagaagaagatggga 20

DB 46222 aaagagaagaagatggga 46240

RESULT 8

AAH05417/c

AAH05417 standard; CDNA; 819 BP.

XX AC

AAH05417;

XX DT

26-JUN-2001 (first entry)

XX DE

Human CDNA clone (5'-primer) SEQ ID NO:2252.

XX KW

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS

Homo sapiens.

XX PN

EP1074617-A2.

XX PD

07-FEB-2001.

XX PF

28-JUL-2000; 2000EP-0116126.

XX PR

29-JUL-1999; 99JP-0248036.

XX PR

27-AUG-1999; 99JP-0300253.

XX PR

11-JAN-2000; 2000JP-0118776.

XX PR

02-MAY-2000; 2000JP-0183767.

XX PR

09-JUN-2000; 2000JP-0241899.

XX PA

(HELI-) HELIX RES INST.

XX PI

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR

WPI; 2001-318749/34.

XX PT

Primer sets for synthesizing polynucleotides, particularly the 5602

XX PT

full-length cDNAs defined in the specification, and for the detection

XX PT

and/or diagnosis of the abnormality of the proteins encoded by the

XX PT

full-length cDNAs -

XX PS

Claim 1; SEQ ID 2252; 2537pp + CD ROM; English.

XX CC

The present invention describes primer sets for synthesizing 5602

XX CC

full-length cDNAs defined in the specification. Where a primer set

XX CC

comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX CC

to the complementary strand of a polynucleotide which comprises one of

XX CC

the 5602 nucleotide sequences defined in the specification, where the

XX CC

oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SQ Sequence 819 BP; 171 A; 286 C; 156 G; 203 T; 3 other;

Query Match 3.9%; Score 18; DB 22; Length 819;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agagaatggagcagcatg 27
|||||
DB 786 AGAGAAATGGAGCAGCATG 769

RESULT 9

AAH14334
ID AAH14334 standard; DNA: 686 BP.

AC AAX14334;

DT 31-MAR-1999 (first entry)

DE H. pylori GHP0 212 gene.

KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.

OS Helicobacter pylori.

FT Key Location/Qualifiers
FT CDS 47..631
FT /tag= a

W09843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-US06371.

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0833457.

PR 24-JUN-1997; 97US-0881227.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

DR WPI: 1998-542293/46.

DR P-PSDB; AAW98615.

PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases

PS Claim 1; Page 1242-1243; 2054pp; English.

XX This sequence represents a polynucleotide of the invention. It was

CC isolated from Helicobacter pylori and encodes a H. pylori GHP0 protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.

SQ Sequence 686 BP; 221 A; 123 C; 161 G; 181 T; 0 other;

Query Match 3.7%; Score 17; DB 19; Length 686;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 tgctgataagattggc 48
|||||
DB 549 tgctgataagattggc 565

RESULT 10

AAA26436
ID AAA26436 standard; cDNA: 1248 BP.

AC AAA26436;

DT 29-JUN-2000 (first entry)

DE Human secreted protein gene 91 SEQ ID NO:101.

KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antiinflammatory; diuretic; neuroprotective; antiallergic;
KW osteoporotic; antiarthritic; antibacterial; antidiabetic; antisthma;
KW antipsoriatic; cardiac; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour; ss.

OS Homo sapiens.

PN W020000698-A1.

PD 10-FEB-2000.

PF 29-JUL-1999; 99WO-US17130.

PR 30-JUL-1998; 98US-0094657.

PR 05-AUG-1998; 98US-0095486.

PR 06-AUG-1998; 98US-0095454.

PR 06-AUG-1998; 98US-0095455.

PR 12-AUG-1998; 98US-0096319.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;

PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;

PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;

DR WPI: 2000-195282/17.

DR P-PSDB; AAY91541.

PT New isolated human genes and the secreted polypeptides they encode,
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders

PS Claim 1; Page 435; 634pp; English.

XX The polynucleotide sequences given in AAA26436 to AAA26458 encode the
XX human secreted proteins given in AAY91451 to AAY91691. The human secreted
XX proteins can have activities based on the tissues and cells they are
XX expressed in. Examples of the activities are: cytostatic;
XX immunosuppressive; antiHIV; antiinflammatory; neurotropic; neuroprotective;
XX antiallergic; osteoporotic; antiarthritic; antibacterial; antidiabetic;
XX antisthma; antipsoriatic; and cardiac. The polynucleotides and their
XX corresponding secreted proteins are useful for preventing, treating or

Sequence 1248 BP; 350 A; 412 C; 238 G; 248 T; 0 other;

Qy	11	gagaatggacacatg	27
Db	-		
	328	gagaatggacacatg	344

ID AAC76612 standard; cDNA; 1278 BP.

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2167 polynucleotide sequence SEQ ID NO:4333.

Human, open reading frame; ORFX detection; cystostatic; hepatotropic; vulnerability; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antiabiotic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antihelminthic; antihypertensive; antineoplastic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

OS Homo sapiens

PN W0200058473-A2

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 05-APR-1999; 99US-0127728.

[illegible]

PI Shimkets RA, Leach M;

XX

PT Novel nucleic acids and peptides derived from open reading frame X,

PT neurodegenerative disorders and cardiovascular disease -

PS Claim 5; Page 3528-3529; 5507pp; English

CC AAC7444.0: AAC7606 encode the proteins given in ABA04237 to ABA43397.
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnereary;
CC antiporatic; antiparkinsonian; neutropic; neuroprotective;
CC osteopathic; anticoagulant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiatic; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antihistamatory; antibacterial; antiviral; antifungal; antipneumatic;
CC antithyroid; and antineuasic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hyperension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergic, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX

Query Match	3.7%	Score 17	DB 21	Length 1278
Best Local Similarity	100.0%	Pred. No. 28		
Matches 17	Conservative	0	Indels	0
			Gaps	0

QY	11	gagaatg	gcacatg	27
Db	1183	gagaatg	gcacacatg	1195

RESULT	12
AAC87123	
ID	AAC87123 standard; cDNA; 1648 BP

AC AAC87123;

DT 20-APR-2001 (first entry)

DE Nucleotide sequence of a plant auxin transporter polypeptide.

KW Auxin; auxin transporter; AUX1; root gravitropism; transgenic plant,

plant phenotype; herbicide; ss

Glycine max. OS
xy

EM	Key	Location/Qualifiers
3	1448	

```

    /*lag= d
    /product= "auxin transporter"

```

XX
XX
PN W0200078965-A2

XX
PD 28-DEC-2000

21-TTN-2000: 2000WO-US17050

AA
PR 22-JUN-1999; 99US-0140212

PA (DUPO) DU PONT DE NEUVOURS & CO E 1
XX

PI Cahoon RE, Weng Z;
XX
XX WPI: 2001-071396/08.
DR P-PSDB: AAB31241.
XX
PT New polynucleotide encoding protein that mediates effects of
PT phytohormone auxin, useful for screening for protein inhibitors for
PT potential as herbicides -
XX
PS Claim 2; Page 55; 69pp; English.
XX
CC The present sequence encodes an auxin transporter. The polypeptide
CC mediates the effects of the plant hormone auxin. The polypeptides of
CC the invention are homologous to the Arabidopsis auxin transporter AUX1,
CC which effects root gravitropism. The polynucleotides are used for
CC creating transgenic plants in which the auxin polypeptides are present
CC at higher or lower levels than normal plants. Plant phenotype can be
CC changed by specifically inhibiting expression of one or more genes by
CC antisense inhibition or cosuppression. The auxin polypeptides can be
CC used as a target to facilitate design and/or identification of inhibitors
CC of AUX1 that may be useful as herbicides. The auxin polynucleotides
CC may also be used as probes for gene mapping information that is useful
CC in plant breeding in order to develop lines with desired phenotypes.
SQ Sequence 1648 BP; 404 A; 365 C; 364 G; 515 T; 0 other;

Query Match 3.7%; Score 17; DB 22; Length 1648;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaaagagaagaagatg 17
|||||
Db 322 gaaagagaagaagatg 338

RESULT 13

AAC87132
ID AAC87122 standard; cDNA; 1775 BP.

AC AAC87122;

DT 20-APR-2001 (first entry)

DE Nucleotide sequence of a plant auxin transporter polypeptide.

XX Auxin; auxin transporter; AUX1; root gravitropism; transgenic plant;
KW plant phenotype; herbicide; ss.

XX Glycine max.

PN WO200078965-A2.

PD 28-DEC-2000.

PF 21-JUN-2000; 2000WO-US17050.

PR 22-JUN-1999; 99US-0140212.

PA (DUPD) DU PONT DE NEMOURS & CO E. I.

PI Cahoon RE, Weng Z;

DR WPI: 2001-071396/08.

PT New polynucleotide encoding protein that mediates effects of
PT phytohormone auxin, useful for screening for protein inhibitors for
PT potential as herbicides -

PS Claim 2; Page 54-55; 69pp; English.

CC The present sequence encodes an auxin transporter. The polypeptide
CC mediates the effects of the plant hormone auxin. The polypeptides of

CC the invention are homologous to the Arabidopsis auxin transporter AUX1,
CC which effects root gravitropism. The polynucleotides are used for
CC creating transgenic plants in which the auxin polypeptides are present
CC at higher or lower levels than normal plants. Plant phenotype can be
CC changed by specifically inhibiting expression of one or more genes by
CC antisense inhibition or cosuppression. The auxin polypeptides can be
CC used as a target to facilitate design and/or identification of inhibitors
CC of AUX1 that may be useful as herbicides. The auxin polynucleotides
CC may also be used as probes for gene mapping information that is useful
CC in plant breeding in order to develop lines with desired phenotypes.
XX
SQ Sequence 1775 BP; 447 A; 390 C; 376 G; 562 T; 0 other;

Query Match 3.7%; Score 17; DB 22; Length 1775;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaaagagaagaagatg 17
|||||
Db 322 gaaagagaagaagatg 338

RESULT 14

AAC44199
ID AAC44199 standard; DNA; 2075 BP.

AC AAC44199;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment smg ID NO: 41988.

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123380.

PR 09-MAR-1999; 99US-0123388.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 3.7%; Score 17; DB 21; Length 2075;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 437 cgatcgcttttgcaaa 453
 |||||
 Db 1632 cgatcgcttttgcaaa 1648

RESULT 15

AAQ79913
 ID AAQ79913 standard; DNA; 3122 BP.

XX AAQ79913;

XX 01-AUG-1995 (first entry)

DE Lys-aminopeptidase pepn gene.

XX Lys-aminopeptidase; pepn gene; fermented food; cheese; ds.

OS Lactobacillus delbrueckii subsp. lactis.

XX key location/Qualifiers

FT CDS 316..2847

XX EP633316-A.

XX 11-JAN-1995.

PD 30-JUN-1994; 94EP-0401497.

XX 01-JUL-1993; 93GB-0013586.

XX (EECE-) EEC EURO ECONOMIC COMMUNITY.

PI Klein JR, Plapp R;

DR WPI; 1995-038513/06.

XX P-PSDB; AAR67760.

PS Claim 8; Page 17-21; 41pp; English.

CC A new Lys-aminopeptidase, pepn, was isolated from *L. delbrueckii*
 CC subsp. *lactis* MS87 (DSM 7290) and had the sequence given in
 CC AAR67760. The pepn gene (AAQ79913) was isolated from a library
 CC of DSM 7290 chromosomal DNA by screening for peptolytic activity
 CC in *Escherichia coli* ER1562 transformants. The isolated gene
 CC is used for recombinant pepn production.

SQ Sequence 3122 BP; 861 A; 745 C; 796 G; 720 T; 0 other;

Query Match 3.7%; Score 17; DB 16; Length 3122;

Best Local Similarity 100.0%; Pred. No. 28;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 180 ctctccaggactcc 196
 |||||
 Db 395 ctctccaggactcc 411

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2002, 00:18:18 ; Search time 81.38 Seconds
(without alignments)
1291.298 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464
Sequence: 1 gaagaagaagaagaatggga.....tttggcaacgaaccgagac 464

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 Residues

Word size: 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	4.1	72928	3	US-09-009-913-1
2	17	3.7	11298	1	US-07-869-933-31
3	17	3.7	11298	1	US-08-201-879A-2
4	17	3.7	11298	4	US-09-103-663-31
5	16	3.4	1724	4	US-08-509-024-1
6	16	3.4	1724	4	US-09-333-279-1
7	16	3.4	1724	5	PCT-US96-12374-1
8	16	3.4	1929	4	US-09-146-950-1
9	16	3.4	2284	2	US-08-467-822-28
10	16	3.4	2284	4	US-08-432-697-28
11	16	3.4	2284	4	US-08-466-248-28
12	16	3.4	2344	3	US-08-893-852A-2
13	16	3.4	3177	1	US-08-042-747A-4
14	16	3.4	3561	3	US-08-822-324-7
15	16	3.4	3561	4	US-08-972-927-4
16	16	3.4	10342	4	US-08-972-927-5
17	16	3.4	15378	3	US-08-785-420-1
18	16	3.4	246240	2	US-08-724-394A-20
19	16	3.4	246240	2	US-08-724-394A-21
20	16	3.4	246240	2	US-08-724-394A-22
21	15	3.2	306	4	US-09-147-935A-10
22	15	3.2	702	3	US-08-842-976-2
23	15	3.2	702	3	US-09-213-397-2
24	15	3.2	702	3	US-09-416-489-2
25	15	3.2	822	4	US-08-896-164-47
26	15	3.2	895	4	US-08-924-747-21
27	15	3.2	895	4	US-09-247-373B-21

28	15	3.2	895	4	US-09-296-715-21	Sequence 21, Appl
29	15	3.2	1058	3	US-09-156-807-1	Sequence 1, Appl
30	15	3.2	1189	3	US-08-961-083-15	Sequence 15, Appl
31	15	3.2	1569	3	US-08-821-984-9	Sequence 9, Appl
32	15	3.2	1569	4	US-09-329-749-9	Sequence 9, Appl
33	15	3.2	1659	4	US-08-548-509-4	Sequence 4, Appl
34	15	3.2	1697	1	US-08-181-271A-104	Sequence 104, App
35	15	3.2	1697	1	US-08-449-315-104	Sequence 104, App
36	15	3.2	1697	1	US-08-444-803-104	Sequence 104, App
37	15	3.2	1697	1	US-08-449-043-104	Sequence 104, App
38	15	3.2	1697	1	US-08-456-265A-104	Sequence 104, App
39	15	3.2	1697	1	US-08-455-416-104	Sequence 104, App
40	15	3.2	1697	1	US-08-455-244-104	Sequence 104, App
41	15	3.2	1697	1	US-08-454-876-104	Sequence 104, App
42	15	3.2	1697	2	US-08-457-364-104	Sequence 104, App
43	15	3.2	1697	2	US-08-456-262-104	Sequence 104, App
44	15	3.2	1697	2	US-08-456-240-104	Sequence 104, App
45	15	3.2	1697	2	US-08-455-736-104	Sequence 104, App

ALIGNMENTS

RESULT 1
US-09-009-913-1
; Sequence 1, Application US/0900913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: Axy's Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SRO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 72928 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-1
Query Match 4.1%; Score 19; DB 3; Length 72928;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 aaagaagaagaatggga 20

Db 46222 AAGAGAAAGAGATGGGA 46240

RESULT 2

US-07-869-933-31
; Sequence 31, Application US/07869933
; Patent No. 5770396

; GENERAL INFORMATION:
; APPLICANT: KINET, Jean-Pierre
; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF
; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR
; TITLE OF INVENTION: IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA

; COUNTRY: USA
; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/869,933

; FILING DATE: 19920416

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 40399/154 NIHD

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109

; -TELEX: 899149

; INFORMATION FOR SEQ ID NO: 31:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11298 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; ORIGINAL SOURCE:

; ORGANISM: homo sapien

; STRAIN: FCRI beta

; US-07-869-933-31

Query Match 3.7%; Score 17; DB 1; Length 11298;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 agagaagaagaatggga 20
|||||

Db 9161 AGAGAAAGAGATGGGA 9177

RESULT 3

US-08-201-879A-2
; Sequence 2, Application US/08201879A

; Patent No. 5807988

; GENERAL INFORMATION:

; APPLICANT: KINET, Jean-Pierre

; APPLICANT: JOUVIN, Marie-Helene

; TITLE OF INVENTION: ISOLATION, CHARACTERIZATION, AND USE OF

; TITLE OF INVENTION: THE HUMAN B SUBUNIT OF THE HIGH AFFINITY RECEPTOR FOR

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/201,879A

; FILING DATE: 24-FEB-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/869,933

; FILING DATE: 16-APR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/03419

; FILING DATE: 16-APR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 40399/234/NIHD

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; -TELEX: 904136

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11298 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; STRAIN: FCRI beta

; FEATURE:

; NAME/KEY: CDS

; LOCATION: join(456..511, 1381..1510, 2026..2160, 4475..4531,

; LOCATION: 5079..5237, 5640..5738, 7224..7319)

; US-08-201-879A-2

Query Match 3.7%; Score 17; DB 1; Length 11298;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 agagaagaagaatggga 20
|||||

Db 9161 AGAGAAAGAGATGGGA 9177

RESULT 4
US-09-103-663-31
; Sequence 31, Application US/09103663D

; Patent No. 6171803

; GENERAL INFORMATION:

; APPLICANT: Kinet et al.

; TITLE OF INVENTION: Isolation, characterization, and use of the human beta

; TITLE OF INVENTION: subunit of the high affinity receptor for

; FILE REFERENCE: 50490

; CURRENT APPLICATION NUMBER: US/09/103,663D

; EARLIER FILING DATE: 1998-06-23

; EARLIER APPLICATION NUMBER: 07/869,933

; NUMBER OF SEQ ID NOS: 35

; SOFTWARE: PatentIn ver. 2.1

; SEQ ID NO 31

; LENGTH: 11298

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-103-663-31

Query Match 3.7%; Score 17; DB 4; Length 11298;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 agagaagaagaatgga 20
|||||
DB 9161 agagaagaagaatgga 9177

RESULT 5
US-08-509-024-1/c
; Sequence 1, Application US/08509024B
; Patent No. 6291207
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/08/509,024B
; CURRENT FILING DATE: 1995-07-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-509-024-1

Query Match 3.4%; Score 16; DB 4; Length 1724;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaat 16
|||||
DB 198 GAAAGAGAAAGAGAAT 183

RESULT 6
US-09-333-279-1/c
; Sequence 1, Application US/09333279
; Patent No. 6303336
; GENERAL INFORMATION:
; APPLICANT: SPEAR, Patricia G.
; APPLICANT: MONTGOMERY, Rebecca I.
; TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
; FILE REFERENCE: 0290-1
; CURRENT APPLICATION NUMBER: US/09/333,279
; CURRENT FILING DATE: 1999-06-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1724
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-333-279-1

Query Match 3.4%; Score 16; DB 4; Length 1724;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaat 16
|||||
DB 198 GAAAGAGAAAGAGAAT 183

RESULT 7
PCT-US96-12374-1/c
; Sequence 1, Application PC/TUS9612374
; GENERAL INFORMATION:

APPLICANT: Northwestern University
; TITLE OF INVENTION: Herpes Virus Entry Mediator
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Goldsmith, Milnamow & Katz, Ltd.
; STREET: 180 N. Stetson, Suite 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/12374
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Northrup, Thomas F.
; REGISTRATION NUMBER: 33,268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5400
; TELEFAX: (312) 616-5460
; TELEX: --
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1724 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 294..1145
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 294..1142
PCT-US96-12374-1

Query Match 3.4%; Score 16; DB 5; Length 1724;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaagagaagaagaat 16
|||||
DB 198 GAAAGAGAAAGAGAAT 183

RESULT 8
US-09-146-950-1/c
; Sequence 1, Application US/09146950A
; Patent No. 6287808
; GENERAL INFORMATION:
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED
; FILE REFERENCE: 09404/057001
; CURRENT APPLICATION NUMBER: US/09/146,950A
; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1929
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (297)...(875)

US-09-146-950-1

Query Match 3.4%; Score 16; DB 4; Length 1929;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaagagagaagaat 16
|||||
Db 195 GAAAGAGAGAAGAAT 180

RESULT 9

US-08-467-822-28
Sequence 28, Application US/08467822
Patent No. 5843460

GENERAL INFORMATION:

APPLICANT: Labigne, Agnes

APPLICANT: Sauerbaum, Sebastien

APPLICANT: Ferrero, Richard L.

APPLICANT: Thiberge, Jean-Michel

TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

POLYPEPTIDES

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Flunegan, Henderson, Farabow, Garrett &

ADDRESSEE: Dunner

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,822

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/447,177

FILING DATE: 19-MAY-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03495.0137-02000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 2284 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-467-822-28

Db 300 GATCGCTTTGGCAAA 315

RESULT 10
US-08-432-697-28
Sequence 28, Application US/08432697
Patent No. 6248330

GENERAL INFORMATION:

APPLICANT: Labigne, Agnes

APPLICANT: Sauerbaum, Sebastien

APPLICANT: Ferrero, Richard L.

APPLICANT: Thiberge, Jean-Michel

TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

POLYPEPTIDES

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Flunegan, Henderson, Farabow, Garrett &

ADDRESSEE: Dunner

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/432,697

FILING DATE: 02-MAY-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146

REFERENCE/DOCKET NUMBER: 03495.0137-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 2284 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-432-697-28

OY 438 gatcgctttggcaaa 453
|||||
Db 300 GATCGCTTTGGCAAA 315

RESULT 11

US-08-466-248-28
Sequence 28, Application US/08466248
Patent No. 6258359

GENERAL INFORMATION:

APPLICANT: Labigne, Agnes

APPLICANT: Sauerbaum, Sebastien

APPLICANT: Ferrero, Richard L.

APPLICANT: Thiberge, Jean-Michel

TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST

TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE

TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID

Query Match 3.4%; Score 16; DB 2; Length 2284;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 438 gatcgctttggcaaa 453

TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,248
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
* TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 2284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-466-248-28

Query Match 3.4%; Score 16; DB 4; Length 2284;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 gatcgcttggcaaa 453
|||||
DB 300 GATCGCTTTGGCAAA 315

RESULT 12
US-08-893-852A-2
Sequence 2, Application US/08893852A
Patent No. 6080558
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: NEW HUMAN GROWTH REGULATOR PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,852A
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0341 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2344 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TMLR3DF01
CLONE: 508302
US-08-893-852A-2

Query Match 3.4%; Score 16; DB 3; Length 2344;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 atgaacgcgtggcct 417
|||||
DB 119 ATGACGCCGTGGCCT 134

RESULT 13
US-08-042-747A-4/C
Sequence 4, Application US/08042747A
Patent No. 5487969
GENERAL INFORMATION:
APPLICANT: Eberle, Richard
APPLICANT: Black, Darla
APPLICANT: Schiacciello, Franco
APPLICANT: Hilliard, Julia K.
TITLE OF INVENTION: Cloning and Amplification of Monkey B
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cox & Smith Incorporated
STREET: 112 East Pecan Street, Suite 2000
CITY: San Antonio
STATE: Texas
COUNTRY: USA
ZIP: 78205
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/042,747A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haymond, W. Bradley
REGISTRATION NUMBER: 35186
REFERENCE/DOCKET NUMBER: S-0072.179
TELECOMMUNICATION INFORMATION:
TELEPHONE: 210-554-5500
TELEFAX: 210-226-8395
TELEX: 767609

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 269..2941
NAME/KEY: CDS
LOCATION: 1..249
US-08-042-747A-4

Query Match 3.4%; Score 16; DB 1; Length 3177;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 368 catcagtcctccgagc 383
|||||
Db 1933 CATCAGTCCCGAGC 1918

RESULT 14
US-08-822-324-7/C
Sequence 7, Application US/08822324
Patent No. 6129917
GENERAL INFORMATION:
APPLICANT: Potempa, Jan S.
APPLICANT: Travis, James
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING
TITLE OF INVENTION: PORPHYROMAN GINGIVALIS PROTEINS AND/OR PEPTIDES AND
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,324
FILING DATE: 21-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,945
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 103-95 WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 488-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3561 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS

LOCATION: 1336..2862
US-08-822-324-7

Query Match 3.4%; Score 16; DB 3; Length 3561;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 tctccaccttctcag 188
|||||
Db 826 TCTCCACCTTCTCAG 811

RESULT 15
US-08-972-927-4
Sequence 4, Application US/08972927
Patent No. 6166290
GENERAL INFORMATION:
APPLICANT: Rea, Phillip A.
APPLICANT: Lu, Yu-Ping
APPLICANT: Li, Ze-Sheng
TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
STREET: Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: US
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,927
FILING DATE: 18-NOV-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,040
FILING DATE: 18-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/061,328
FILING DATE: 08-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-1202
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831494
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5175 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-972-927-4

Query Match 3.4%; Score 16; DB 4; Length 5175;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 149 gagaattcagtggtg 164
|||||
Db 4239 GAGAAATTCAGTGTG 4254

'Sat Apr 6 11:03:27 2002

us-09-621-781-1_copy_1507_1970.rni

Page 7

Search completed: April 6, 2002, 06:18:07
Job time: 21589 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2002, 00:07:08 ; Search time 1415.1 Seconds
(without alignments)
3523.453 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464
Sequence: 1 gaagagagaagaagatgga.....ttggcaacgaaccgac 464

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estipl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_vit:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	7.5	461	10	AA920479
2	35	7.5	542	10	AA762360
3	35	7.5	574	11	BG100299
4	35	7.5	575	11	AA472295
5	35	7.5	596	10	BE573255
6	35	7.5	615	11	W62865
7	35	7.5	623	11	AA320574
8	35	7.5	708	11	BE573350
9	35	7.5	908	11	BF182426
10	35	7.5	1136	10	BE573733
11	26	5.6	446	10	AA684447
12	26	5.6	573	10	AA104251

13	21	4.5	1077	10	BE614567	BE614567	601503841
14	20	4.3	454	10	AA634278	AA634278	b118901.w
15	20	4.3	469	10	AA634288	AA634288	b118901.w
16	20	4.3	471	10	AA789164	AA789164	ag58b01.s
17	20	4.3	492	13	AO536955	AO536955	RPCT-11-3
18	20	4.3	515	10	AT759543	AT759543	ETESTea29
19	20	4.3	594	13	AO537404	AO537404	RPCT-11-3
20	20	4.3	936	11	B1113465	B1113465	602899696
21	19	4.1	191	10	AV069346	AV069346	AV069346
22	19	4.1	230	10	BB017230	BB017230	BB017230
23	19	4.1	249	10	BB252955	BB252955	BB252955
24	19	4.1	330	10	AV212169	AV212169	AV212169
25	19	4.1	400	11	BF715432	BF715432	BF715432
26	19	4.1	432	13	A2817262	A2817262	2M0086003
27	19	4.1	540	11	BC363063	BC363063	sac099056
28	19	4.1	632	13	A2937634	A2937634	2M0195F22
29	19	4.1	652	13	A2949846	A2949846	2M0213H17
30	19	4.1	746	10	BE038724	BE038724	AB04F03.A
31	19	4.1	836	10	BE782858	BE782858	601472368
32	18	3.9	79	10	A1966441	A1966441	sc39b06.y
33	18	3.9	150	13	BH055024	BH055024	RPCT-24-2
34	18	3.9	173	13	AO389139	AO389139	RPCT11-15
35	18	3.9	228	11	BF716107	BF716107	saal4f03
36	18	3.9	229	10	A1506763	A1506763	vm58c08.x
37	18	3.9	248	10	AA596509	AA596509	vm58c08.t
38	18	3.9	253	13	A2621106	A2621106	1M0454D08
39	18	3.9	262	10	BE058354	BE058354	sn14h1.y
40	18	3.9	263	11	BG049447	BG049447	CV1-20.A1
41	18	3.9	280	10	AM134889	AM134889	UI-H-B11
42	18	3.9	363	10	BE363395	BE363395	MS1.62.D1
43	18	3.9	364	13	A2553147	A2553147	RPCT-23-2
44	18	3.9	374	13	A2655779	A2655779	1M0530L20
45	18	3.9	378	10	A1494185	A1494185	1149g10.y

ALIGNMENTS

RESULT 1
LOCUS AA920479
DEFINITION AA920479 461 bp mRNA
IMAGE:129182 5' similar to gb:51829 Mouse myeloid differentiation primary response mRNA encoding (MUSE);, mRNA sequence.

ACCESSION AA920479
VERSION AA920479.1 GI:3067258
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 461)
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenger, K., Stepien, M., Tan, F., Underwood, K., Moore, B., Theisinger, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:680230
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 348.
Location/Qualifiers
1..461

```

/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:1299182"
/clone_lib="Stratagene mouse lung 937302"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 month old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: plusscript SK-; Site: 1; EcorI
; Site: 2; XhoI: Cloned unidirectionally. Primer: Oligo
dr. 6-8 month old female lung and 1.5 year old male lung
were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP
XR Vector; -5' adaptor sequence: 5' GAATTCGACACGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"
BASE COUNT      102 a      157 c      125 g      77 t
ORIGIN

Query Match      7.5%: Score 35; DB 10; Length 461;
Best Local Similarity 100.0%; Pred. No. 3.1e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      373  cgtcccgagcccgccctgagggcgacatgaac 407
            |||||||
Db      66  CGTCCCGAGCCCGCCTGAGGCGACATGAC 100

RESULT      2
AM762360      542 bp      mRNA      EST      04-MAY-2000
LOCUS
DEFINITION    uc56h03.y1 NCI-CGAP_Mam3 Mus musculus cDNA clone IMAGE:3154325 5'
              similar to gb:X51829 Mouse myeloid differentiation primary response
              mRNA encoding (MOUSE);, mRNA sequence.
ACCESSION     AM762360.1 GI:7694285
VERSION
KEYWORDS
SOURCE
ORGANISM      house mouse.
              Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 542)
REFERENCE     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
              Unpublished (1997)
JOURNAL
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Other ESTs: uc56h03.x1
              Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              image.llnl.gov/image/html/resources.shtml
MGI:1057081
Seq primer: -40RP from Gibco
High quality sequence stop: 374.
Location/Qualifiers
1..542
/organism="Mus musculus"
/strain="129 - C57/B6 - FVB"
/db_xref="taxon:10090"
/clone="IMAGE:3154325"
/clone_lib="NCI-CGAP Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site: 1; SalI;
              Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dr.
              library constructed by Life Technologies. Investigators

```

```

/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:3655420"
/clone_lib="McCarrey Eddy type B spermatogonia"
/sex="male"
/tissue_type="type B spermatogonia, pooled from multiple
mice"
/dev_stage="8 day"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pluscript SK+ (Stratagene
); Site: 1; XhoI; Site: 2; EcorI; cDNA oligo dr-primed
[5'-(GA)10-ACTAGCTCGAGTTTCTTTTCTTTT-3'] and directionally
cloned using 5' linkers 5'-AATTCGACGAG-3' and
5'-CTCGTCGCG-3'. Size selection of >400bp material gives
average insert size ranging from 1-2 kb. Library was mass
excised (from lambda-uniZAP-XR) and resulting
single-stranded phagemids were prepped and transformed
into DH10B. Library contains 96% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National

```

BASE COUNT

114 a 171 c 159 g 98 t

ORIGIN

Query Match

7.5%: Score 35; DB 10; Length 542;

Best Local Similarity

100.0%; Pred. No. 3.2e-07;

Matches 35; Conservative

0; Mismatches 0; Indels 0; Gaps 0;

QY

373

cgtcccgagcccgcccgctgagggcgacatgaac 407

Db

87

CGTCCCGAGCCCGCCTGAGGCGACATGAC 121

RESULT 3

BG100299

LOCUS

DEFINITION

ucx84f03.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA

clone IMAGE:3655420 5' similar to SW:MY16_MOUSE P17564 MYELOID

DIFFERENTIATION PRIMARY RESPONSE PROTEIN MYD116. ;, mRNA sequence.

ACCESSION

BG100299

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 574)

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person

, E., Kohn, S., Shin, T., Jackson, T., Cardenas, M., McCann, R.,

Waterson, R., and Wilson, R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:1417724

Seq primer: Primer name ambiguous

High quality sequence stop: 392.

Location/Qualifiers

1..574

/organism="Mus musculus"

/strain="CD-1"

/db_xref="taxon:10090"

/clone="IMAGE:3655420"

/clone_lib="McCarrey Eddy type B spermatogonia"

/sex="male"

/tissue_type="type B spermatogonia, pooled from multiple

mice"

/dev_stage="8 day"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis; Vector: pluscript SK+ (Stratagene

); Site: 1; XhoI; Site: 2; EcorI; cDNA oligo dr-primed

[5'-(GA)10-ACTAGCTCGAGTTTCTTTTCTTTT-3'] and directionally

cloned using 5' linkers 5'-AATTCGACGAG-3' and

5'-CTCGTCGCG-3'. Size selection of >400bp material gives

average insert size ranging from 1-2 kb. Library was mass

excised (from lambda-uniZAP-XR) and resulting

single-stranded phagemids were prepped and transformed

into DH10B. Library contains 96% recombinants.

References: J. Androl. 20:635-639 and Gene 25:263-269.

Library constructed and donated by J. McCarrey, Ph.D.

(Southwest Foundation for Biomedical Research, Dept. of

Genetics); excision done by E.M. Eddy, Ph.D. (National

Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #33417."

Query Match 7.5%; Score 35; DB 11; Length 574;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 373 cgtcccgagccccacgctttagggcagcatgtaac 407
|||||
Db 51 CGTCCCGAGCCCCACGCTTAGGGCGACATGAAC 85

RESULT	4
AA472295	575 bp mRNA EST 18-JUN-1997
LOCUS	
DEFINITION	VH01C05.r1 Soares_mammary_gland_NbMNG Mus musculus cDNA clone

ACCESSION	AA472295
VERSION	AA472295.1
KEYWORDS	EST.
SOURCE	house mouse.

SOURCE	ORGANISM	house mouse. Mus musculus
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 575)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouesest@watson.wustl.edu
This clone is available royalty-free through LNC ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
GSI:513664

FEATURES
source

```
1. .575
/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clonimg="IMAGE:874184"
/clonimg_1bp="Soares,mammary_gland_nbmg"
/sex="male"
/tissue_type="mammary_gland"
/dev_stage="4 weeks"
```

BASE COUNT	127 a	183 c	161 g	104 f
ORIGIN				

Query Match	7.5%;	Score 35;	DB 10;	Length 575;
Best Local Similarity	100.0%;	Pred. No. 3.2e-07;		
Matches 35;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy 373 cgtcccgagccccacgacctgaaggcgacatgaaac 407
 |||
 Db 93 CGTCCCGAGCCCCACGCGCTGAGGGCGACATGAAC 127

RESULT	5			
BE573255				
LOCUS	BE573255	596 bp	mRNA	EST
DEFINITION	601333152P2	NCI-CGAP_Mam6	Mus musculus	CDNA clone IMAGE:3710629
	mRNA sequence.			5' ,

ACCESSION	BE573255
VERSION	BE573255.1
KEYWORDS	house mouse, EST.
SOURCE	Mus musculus
ORGANISM	

SOURCE	ORGANISM	Mus musculus	Mus musculus
...

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
1 (bases 1 to 596)
TITLE NIH-MGC <http://mgc.ncl.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Jeffrey Green M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLM6933 row: g column: 14
 High quality sequence seq: 542.

FEATURES

SOURCE

```

1..596
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3710629"
/clone_lib="NCI CGAP Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Sa
Site:2: NotI; Cloned unidirectionally. Primer: Oligo
library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
a      199 c      163 g      102 t

```

BASE COUNT
ORIGIN

```
Query Match      7.5%; Score 35; DB 10; Length 596;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
```

```

QY      373  cgtcccgagcccaagcctgagggcagcatgaac 407
          |||||
Db      82  CGTCCCGAGCCCCACGCGCTGAGGGGACATGAAC 116

```

RESULT	6						
W62865		615 bp	mRNA			EST	07-JUN-1996
LOCUS	W62865						
DEFINITION	md65f08.r1 Soares mouse embryo		NBMEL3.5	14.5	Mus musculus	CDNA	

differentiation primary response mRNA encoding (MOUSE); mRNA sequence.

ACCESSION W62865
 VERSION W62865.1 GI:1369614
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 615)
 REFERENCE 1 (bases 1 to 615)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellander, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:236639
 Seq primer: ETPPrimer
 High quality sequence stop: 344.
 Location/Qualifiers
 1..615
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:375207"
 /clone_lib="Soares mouse embryo NbME13.5 14.5"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCACCAATCTGAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 1; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 145 a 190 c 165 g 115 t.
 ORIGIN

Query Match 7.5%; Score 35; DB 11; Length 615;
 Best Local Similarity 100.0%; Pred. No. 3.2e-07;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 373 cgtcccccagagccacgcctgagggcagacatgaac 407
 |||||||||||||||||||||||||||||||||||||||
 Db 92 cgtccccgagccccacgcctgagggcagacatgaac 126

RESULT 7
 AM320574 623 bp mRNA EST 25-JAN-2000
 LOCUS u020f09.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:2609705 5',
 DEFINITION similar to gb:XS1829 Mouse myeloid differentiation primary response
 mRNA encoding (MOUSE);, mRNA sequence.
 ACCESSION AM320574.1 GI:6750118
 VERSION EST.
 KEYWORDS house mouse.
 SOURCE

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 623)
 REFERENCE 1 (bases 1 to 623)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/db/ftp/image/image.html
 MGI:1023565
 Seq primer: -40RP from Gibco
 High quality sequence stop: 434.
 Location/Qualifiers
 1..623
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:2609705"
 /clone_lib="NCI_CGAP_Mam6"
 /sex="female, virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 137 a 197 c 173 g 115 t
 ORIGIN

Query Match 7.5%; Score 35; DB 10; Length 623;
 Best Local Similarity 100.0%; Pred. No. 3.2e-07;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 373 cgtcccccagagccacgcctgagggcagacatgaac 407
 |||||||||||||||||||||||||||||||||||||||
 Db 119 cgtccccgagccccacgcctgagggcagacatgaac 153

RESULT 8
 BE573350 708 bp mRNA EST 15-AUG-2000
 LOCUS 601333015P2 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3710762 5',
 DEFINITION mRNA sequence.
 ACCESSION BE573350
 VERSION BE573350.1 GI:9817070
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 708)
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov
Plate: LLAM8933 row: m column: 03
High quality sequence stop: 649.

FEATURES

source

Location/Qualifiers
1. 708

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3710762"
/clone_1ib="NCL_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT

165 a 213 c 208 g 122 t

ORIGIN

Query Match 7.5%: Score 35; DB 10; Length 708;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 cgtcccgagcccgctgagggcgacatgaac 407
|||||
Db 73 cgtcccgagcccgctgagggcgacatgaac 107

RESULT 9

LOCUS BF182426 908 bp mRNA EST 31-OCT-2000
DEFINITION 60180418/F1 NCL_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035261 5',
mRNA sequence.
ACCESSION BF182426
VERSION BF182426.1 GI:11060569
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE

1 (bases 1 to 908)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA library preparation: Life Technologies, Inc.
cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM9308 row: m column: 22
High quality sequence stop: 646.

FEATURES

source

Location/Qualifiers
1. 908

/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:4035261"
/clone_1ib="NCL_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

BASE COUNT 220 a 271 c 264 g 153 t
ORIGIN

Query Match 7.5%: Score 35; DB 11; Length 908;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 cgtcccgagcccgctgagggcgacatgaac 407
|||||
Db 92 cgtcccgagcccgctgagggcgacatgaac 126

RESULT 10

LOCUS BE573733 1136 bp mRNA EST 15-AUG-2000
DEFINITION 601333482F2 NCL_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3710924 5',
mRNA sequence.
ACCESSION BE573733
VERSION BE573733.1 GI:9817453
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE

1 (bases 1 to 1136)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue procurement: Jeffrey Green M.D.
cDNA library preparation: Life Technologies, Inc.
cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM8934 row: c column: 21
High quality sequence stop: 635.

FEATURES

source

Location/Qualifiers
1. 1136

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3710924"
/clone_1ib="NCL_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 300 a 363 c 311 g 161 t

ORIGIN

Query Match 7.5%: Score 35; DB 10; Length 1136;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 373 cgtcccgagcccgctgagggcgacatgaac 407
|||||
Db 81 cgtcccgagcccgctgagggcgacatgaac 115

RESULT 11

LOCUS AA684447/c 446 bp mRNA EST 09-DEC-1997
DEFINITION ym64b07.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
IMAGE:1002997 5' similar to gb:x51829 Mouse myeloid differentiation

primary response mRNA encoding (MOUSE);, mRNA sequence.
AA684447
AA684447.1 GI:2671033
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 446)
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:567213
High quality sequence stop: 280.
Location/Qualifiers
1..446
/organism="Mus musculus"
/strain="B6D F1/J"
/db_xref="taxon:10090"
/clone_image="IMAGE:1002997"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/note="Organ: embryo; Vector: Bluescribe (modified);
Site_1: Muir; Site_2: Sail; Cloned unidirectionally from
mRNA prepared from 13,500 2-cell stage embryos. Primer:
Sail(5')-5'-CGGCGACGCGTCGCGCTTTTGT-3', CDNAS
were cloned into the Muir/Sail sites of a modified
Bluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."
BASE COUNT 79 a 123 c 154 g 90 t
ORIGIN
Query Match 5.6%; Score 26; DB 10; Length 446;
Best Local Similarity 100.0%; Pred.No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 427 tcggaccacagatcgctttggcaa 452
|||||
Db 294 TCCGACCCACGATCGCTTTGGCAA 269
RESULT 12
AA104251 573 bp mRNA EST 29-OCT-1996
LOCUS AA104251
DEFINITION CDNA clone IMAGE:556973 5' similar to gb:U51829 Mouse myeloid
differentiation primary response mRNA encoding (MOUSE);, mRNA
sequence.
AA104251
VERSION AA104251.1 GI:1650409
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 573)
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev1 from Amersham
MGI:337765
High quality sequence stop: 246.
Location/Qualifiers
1..573
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:556973"
/clone_lib="Life Tech mouse embryo 10 5dpc 10665016"
/tissue_type="embryo"
/dev_stage="10.5dpc embryos"
/lab_host="DH10B"
/note="Organ: whole embryo; Vector: pCMV-SPORT2; Site_1:
Sail; Site_2: Not; Cloned unidirectionally. Primer:
Oligo dt. 10.5dpc embryos. pCMV-SPORT2 vector."
BASE COUNT 123 a 178 c 156 g 115 t 1 others
ORIGIN
Query Match 5.6%; Score 26; DB 10; Length 573;
Best Local Similarity 100.0%; Pred.No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 427 tcggaccacagatcgctttggcaa 452
|||||
Db 147 TCCGACCCACGATCGCTTTGGCAA 172
RESULT 13
BE614567 1077 bp mRNA EST 20-OCT-2000
LOCUS BE614567
DEFINITION 60150384171 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3905554 3',
mRNA sequence.
BE614567
VERSION BE614567.1 GI:9896164
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1077)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L14M9713 row: a column: 11
High quality sequence start: 60
High quality sequence stop: 682.
Location/Qualifiers

source

1. .1077
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3905554"
/clone_11b="NIH-MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

BASE COUNT 289 a 285 c 268 g 235 t

ORIGIN

Query Match 4.5%; Score 21; DB 10; Length 1077;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 acagcatgtactgcctgatal 40
|||||

Db 761 ACAGCATGTACTGCCTGATG 781

RESULT 14
AM634278 454 bp mRNA EST 26-APR-2001
LOCUS b118g01.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
laevis cDNA clone PBX0018G01 5', mRNA sequence.
AM634278
VERSION AM634278.1 GI:7391359
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 454)
Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.
Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman
J.W., Bonaldo, M.F. and Soares, M.B.
The NIH's Xenopus maternal EST project: Interim analysis of the
first 13,879 ESTs from unfertilized eggs
21211403
Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
Phone 800-533-4363 ext.cdn, fax 256-536-9016 att:cdn, email
cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGCGCCAGT
BACKWARD: CAGGAACAGCTATGACG
Plate: 0018 row: G column: 01
Seq primer: T7 primer.

FEATURES
Location/Qualifiers
1. .454
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0018G01"
/clone_11b="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"

/dev_stage="unfertilized egg"
/lab_host="DH10B"
/note="Vector: pT73-Pac; Site: 1; EcoRI; Site: 2; NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT73-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-drl8 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT73-Pac vector.
The library contained approximately 7.2 x 10⁵
recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 130 a 69 c 134 g 121 t

ORIGIN

Query Match 4.3%; Score 20; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 agagaagaagatggagacg 23
|||||

Db 78 AGAGAAGAAGATGGGACAG 97

RESULT 15
AM634288 469 bp mRNA EST 26-APR-2001
LOCUS b118h01.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
laevis cDNA clone PBX0018H01 5', mRNA sequence.
AM634288
VERSION AM634288.1 GI:7391369
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 469)
Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.
Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman
J.W., Bonaldo, M.F. and Soares, M.B.
The NIH's Xenopus maternal EST project: Interim analysis of the
first 13,879 ESTs from unfertilized eggs
21211403
Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
Phone 800-533-4363 ext.cdn, fax 256-536-9016 att:cdn, email
cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGCGCCAGT
BACKWARD: CAGGAACAGCTATGACG
Plate: 0018 row: H column: 01
Seq primer: T7 primer.

FEATURES
Location/Qualifiers
1. .469
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0018H01"
/clone_11b="Blackshear/Soares normalized Xenopus egg
library"

```

library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT73-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dt18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT73-Pac vector.
The library contained approximately 7.2 X 10^5
recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      134 a      79 c      136 g      120 t
ORIGIN

```

```

Query Match      4.3%; Score 20; DB 10; Length 469;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      4 agagaaagagaatggacag 23
      ||||||||||||||||
Db      88 AGAGAAAGAGATGGACAG 107

```

Search completed: April 6, 2002, 05:51:30
Job time: 20662 sec

Matches 122; Conservative 0; Mismatches 48; Indels 2; Gaps 1;

QY 293 gtgagaattatcagagaccccgagctccgcatcccttgcggagacgcttg 352

DB 8 GTTGTGATCTTATCAGACGCTGCACGACCCGCCGCTTGTGCCACGGGACTTG 67

QY 353 cgagagcccgagacatcaagt--ccccgagcccaagcgtgagggcgagatgaacgcg 410

DB 68 AGGACGCGGAGATACCTGTGAGTTACTGAGAGCCCGAGCCTGAGGGATGAGACGCG 127

QY 411 ctgagccttgagagacatccgagcccgagcttgcggagacgacgcg 462

DB 128 CTGGCTCCCTAACCGTCCGACCTGTGATGCTTCTGCGACCGACGCG 179

RESULT 2

US-08-592-874-1

; Sequence 1, Application US/08592874

; Patent No. 5854034

; GENERAL INFORMATION:

; APPLICANT: POLLOCK, THOMAS J.

; APPLICANT: YAMAZAKI, MOTOHIDE

; APPLICANT: THORNE, LINDA

; APPLICANT: MIKOLAJCZAK, MARCIA

; APPLICANT: ARMENTROUT, RICHARD W.

; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING

; NUMBER OF INVENTION: POLYSACCHARIDE PRODUCTION

; CORRESPONDENCE ADDRESS: 1

; ADDRESSEE: JULES E. GOLDBERG

; STREET: 261 MADISON AVENUE

; CITY: NEW YORK

; STATE: NY

; COUNTRY: USA

; ZIP: 10016-2391

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/592,874

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/377,440

; FILING DATE: 24-JAN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: GOLDBERG, JULES E.

; REGISTRATION NUMBER: 24,408

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-986-4090

; TELEFAX: 212-818-9479

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 28804 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; FRAGMENT TYPE: N-terminal

US-08-592-874-1

Query Match 7.3%; Score 33.8; DB 2; Length 28804;

Best Local Similarity 51.0%; Pred. No. 1;

Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 301 ttacatgcgagaccccgagctccgcatcccttgcggagacgcttgcgacgc 360

DB 24702 ttacccgagccgagcttcattcgcctgcgacgtgacccacccgttcgacgacgcg 24761

QY 361 cgtgagacatcaatcccgagcccaagcgcctgagggcgacatgaaagcgctgcttga 420

DB 24762 CGCGGCTCATCATCTCGACACGCTCACCATTATCGGGCAACCGGCTGTGACCGCGA 24821

QY 421 gagcaatccgagccacgcagctcgttgcgaaacgca 457

DB 24822 TCGAGAAAGCCGCCCACTACCGCTTCGTCACCGCGA 24858

RESULT 3

US-09-096-942-2

; Sequence 2, Application US/09096942

; Patent No. 6027925

; GENERAL INFORMATION:

; APPLICANT: Pollock, Thomas J

; APPLICANT: Mikolajczak, Marcia

; APPLICANT: Yamazaki, Motohide

; APPLICANT: Thorne, Linda

; APPLICANT: Armentrout, Richard W

; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria

; FILE REFERENCE: seq list for appl filed from pro. appl

; CURRENT APPLICATION NUMBER: US/09/096,942

; EARLIER FILING DATE: 1998-06-12

; EARLIER APPLICATION NUMBER: 60/049,428

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 2

; LENGTH: 28804

; TYPE: DNA

; ORGANISM: Sphingomonas sp. S88

US-09-096-942-2

Query Match 7.3%; Score 33.8; DB 3; Length 28804;

Best Local Similarity 51.0%; Pred. No. 1;

Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 301 ttacatgcgagaccccgagctccgcatcccttgcggagacgcttgcgacgc 360

DB 24702 ttacccgagccgagcttcattcgcctgcgacgtgacccacccgttcgacgacgcg 24761

QY 361 cgtgagacatcaatcccgagcccaagcgcctgagggcgacatgaaagcgctgcttga 420

DB 24762 cgcgcgtcatactcgcagaaagctcacctatgacgggaacccgctcgtgacgcgca 24821

QY 421 gagcaatccgagccacgcagctcgttgcgaaacgca 457

DB 24822 tcgagaaagcccccactacatcgcgttcgacgcgca 24858

RESULT 4

US-09-096-867-2

; Sequence 2, Application US/09096867

; Patent No. 6030817

; GENERAL INFORMATION:

; APPLICANT: Pollock, Thomas J

; APPLICANT: Mikolajczak, Marcia

; APPLICANT: Yamazaki, Motohide

; APPLICANT: Thorne, Linda

; APPLICANT: Armentrout, Richard W

; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria

; FILE REFERENCE: seq list for appl filed from pro. appl

; CURRENT APPLICATION NUMBER: US/09/096,867

; EARLIER FILING DATE: 1998-06-11

; EARLIER APPLICATION NUMBER: 60/049,428

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 2

; LENGTH: 28804

; TYPE: DNA

RESULT 7
US-08-675-153-3/c
Sequence 3, Application US/08675153
Patent No. 5677124
GENERAL INFORMATION:
APPLICANT: DUBOIS, Dwight
APPLICANT: PASTOSKE, Brittan L.
TITLE OF INVENTION: RIBONUCLEASE RESISTANT VIRAL
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/675,153
FILING DATE: Concurrently Herewith
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: AMB1:026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1718 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-675-153-3

Query Match 6.6%; Score 30.8; DB 1; Length 1718;
Best Local Similarity 48.3%; Pred. No. 2.8;
Matches 86; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 136 ccgaatctccgcgagaaattcagtggtttctctctccacattctcaggaattc 195
1495 CAGGAAGCTCTACACCAACAGCTGCTGCTTACGACCTTGGCATTTGATGG 1436

QY 196 cgaactcgcctctccgtagcgatagcgatgcgctgcacataaactccggg 255
DB 1435 TGTATTGGGATTCGCGAGAGCTCTACAGACGCTACAGGTTACTTGTAAACCTGTG 1376

QY 256 tgatcgtgtggcgagattgactcagtcgagctgtggaagaatlaacatgcgagac 313
DB 1375 AACGCGAGTTAGAGCTGATTCATTGACGAGACCCCGTTAGCGAAGTGTGGGGCGAC 1318

RESULT 8
US-08-841-252-3/c
Sequence 3, Application US/08841252
Patent No. 5919625
GENERAL INFORMATION:
APPLICANT: DUBOIS, DWIGHT
APPLICANT: WINKLER, MATTHEW
APPLICANT: PASTOSKE, BRITTAN L.
TITLE OF INVENTION: RIBONUCLEASE RESISTANT VIRAL RNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD WHITE & DURKEE

STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/841,252
FILING DATE: 29-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 5,677,124
FILING DATE: 03-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARK B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: AMB1:026--1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-300
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1718 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-841-252-3

Query Match 6.6%; Score 30.8; DB 2; Length 1718;
Best Local Similarity 48.3%; Pred. No. 2.8;
Matches 86; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 136 ccgaatctccgcgagaaattcagtggtttctctctccacattctcaggaattc 195
DB 1495 CAGGAAGCTCTACACCAACAGCTGCTGCTTACGACCTTGGCATTTGATGG 1436

QY 196 cgaactcgcctctccgtagcgatagcgatgcgctgcacataaactccggg 255
DB 1435 TGTATTGGGATTCGCGAGAGCTCTACAGACGCTACAGGTTACTTGTAAACCTGTG 1376

QY 256 tgatcgtgtggcgagattgactcagtcgagctgtggaagaatlaacatgcgagac 313
DB 1375 AACGCGAGTTAGAGCTGATTCATTGACGAGACCCCGTTAGCGAAGTGTGGGGCGAC 1318

RESULT 9
US-08-881-571-3/c
Sequence 3, Application US/08881571
Patent No. 5939262
GENERAL INFORMATION:
APPLICANT: PASTOSKE, Brittan L.
APPLICANT: DUBOIS, Dwight
APPLICANT: Brown, David
APPLICANT: Winkler, Matthew
TITLE OF INVENTION: RIBONUCLEASE RESISTANT RNA PREPARATION
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/881,571
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/675,153
FILING DATE: 03-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/021,145
FILING DATE: 03-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: AMBI:033
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1718 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-881-571-3

Query Match
Best Local Similarity 48.3%; Pred. No. 2.8;
Matches 86; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 136 ccgaatctccgcgagaatttcagtggtttctctccacccttcaggagcttc 195
DB 1495 CAGGAAGCTCTACACCAACAGCTGTGGTGCCACTTAAAGCACCCTTGATGG 1436
QY 196 cgaactccgcctccgcgtgacgcagcatagcgcgtgcagactataaactcccg 255
DB 1435 TGTATTTCGATCTTCGCCAGAGCTCTGACGAAGCTTACTTTGTAAGCTGTG 1376
QY 256 tgatcgltgtgycgacattgactgacgttcgcagcttggtgaagattacatgcgagac 313
DB 1375 AACGCGAGTTAGAGCTGATTCATTCAGCGACCCGTTAGCGAAGTTGTTGGCGCGAC 1318

RESULT 10
US-09-282-054-3/c
Sequence 3, Application US/09282054
Patent No. 6214982
GENERAL INFORMATION:
APPLICANT: Pasloske, Brittan L.
APPLICANT: Dubois, Dwight
APPLICANT: Brown, David
APPLICANT: Winkler, Matthew
TITLE OF INVENTION: RIBONUCLEASE RESISTANT RNA PREPARATION
TITLE OF INVENTION: AND UTILIZATION
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/282,054
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/881,571
FILING DATE:
APPLICATION NUMBER: US 08/675,153
FILING DATE: 03-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/021,145
FILING DATE: 03-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: AMBI:033
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1718 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-282-054-3

Query Match
Best Local Similarity 48.3%; Pred. No. 2.8;
Matches 86; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 136 ccgaatctccgcgagaatttcagtggtttctctccacccttcaggagcttc 195
DB 1495 CAGGAAGCTCTACACCAACAGCTGTGGTGCCACTTAAAGCACCCTTGATGG 1436
QY 196 cgaactccgcctccgcgtgacgcagcatagcgcgtgcagactataaactcccg 255
DB 1435 TGTATTTCGATCTTCGCCAGAGCTCTGACGAAGCTTACTTTGTAAGCTGTG 1376
QY 256 tgatcgltgtgycgacattgactgacgttcgcagcttggtgaagattacatgcgagac 313
DB 1375 AACGCGAGTTAGAGCTGATTCATTCAGCGACCCGTTAGCGAAGTTGTTGGCGCGAC 1318

RESULT 11
US-08-675-153-6/c
Sequence 6, Application US/08675153
Patent No. 5677124
GENERAL INFORMATION:
APPLICANT: Dubois, Dwight
APPLICANT: Winkler, Matthew
APPLICANT: Pasloske, Brittan L.
TITLE OF INVENTION: RIBONUCLEASE RESISTANT VIRAL
TITLE OF INVENTION: RNA STANDARDS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/675,153
FILING DATE: Concurrently Herewith
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: AMBI:026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000

```


Db 1495 CAGGAAGCTTACACCAACAGTCTGGTTCACCTTGAAGCACTGCAGCTTGTATG 1436
Qy 196 cgaactcgccttcgcggtgacgtcagatagcgtcgtcagactataactccggg 255
Db 1435 TGTATTTGGATTCGTCCAGAGCTCTGACGACCTACAGTTCCTTGTAAAGCTGTG 1376
Qy 256 tgatcgtgttgccgagatgactcagcttgtagaagatcatalgagagac 313
Db 1375 AACGGAGTTAGAGCTGATTCATTCACGAGACCCCTTACGCAAGTTGCTTGCGGCGAC 1318

RESULT 14

US-09-282-054-6/C
; Sequence 6, Application US/09282054
; Patent No. 6214982
; GENERAL INFORMATION:
; APPLICANT: Pasloske, Brittan L.
; APPLICANT: Dubois, Dwight
; APPLICANT: Brown, David
; APPLICANT: Winkler, Matthew
; TITLE OF INVENTION: RIBONUCLEASE RESISTANT RNA PREPARATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/282,054
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/881,571
; FILING DATE:
; APPLICATION NUMBER: US 08/675,153
; FILING DATE: 03-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/021,145
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: AMI:033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-282-054-6

Query Match

Best Local Similarity 6.6%; Score 30.8; DB 4; Length 1867;
Pred. No. 2.9;

Matches 86; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 136 ccggaatcgcgcgagaattcagtggtttctctctcctcaacttctcaggaactc 195
Db 1495 CAGGAAGCTTACACCAACAGTCTGGTTCACCTTGAAGCACTGCAGCTTGTATG 1436
Qy 196 cgaactcgccttcgcggtgacgtcagatagcgtcgtcagactataaactccggg 255

Db 1435 TGTATTTGCATTCCTGCGCAGAGCTCTGACGAACGCTACAGTTCCTTGTAAAGCTGTG 1376
Qy 256 tgatcgtgttgccgagatgactcagcttgtagaagatcatalgagagac 313
Db 1375 AACGGAGTTAGAGCTGATTCATTCACGAGACCCCTTACGCAAGTTGCTTGCGGCGAC 1318

RESULT 15

US-08-592-874-1/C
; Sequence 1, Application US/08592874
; Patent No. 5834034
; GENERAL INFORMATION:
; APPLICANT: POLLOCK, THOMAS J.
; APPLICANT: YAMAZAKI, MOTOHIDE
; APPLICANT: THORNE, LINDA
; APPLICANT: MIKOLAJCZAK, MARCIA
; APPLICANT: ARMENTROUT, RICHARD W.
; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JULES E. GOLDBERG
; STREET: 261 MADISON AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,874
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/377,440
; FILING DATE: 24-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDBERG, JULES E.
; REGISTRATION NUMBER: 24,408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-986-4090
; TELEFAX: 212-818-9479
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; US-08-592-874-1

Query Match

Best Local Similarity 6.6%; Score 30.6; DB 2; Length 28804;
Pred. No. 10;

Matches 81; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 249 tcccggtgacgtgtgtggtgcagatgactcagcttgcaagcttgtagaagattacatgc 308
Db 28177 TGCCGGCGGTGGCTGCCCAAGATCAAGCGCCGAGACAGCTGGCGAAGATGGCACG 28118
Qy 309 gagaccgcgcgagatcccatcttgcgggagacagccttgcagaccgcgtgagac 368
Db 28117 GTGAAGCTGAAGGTGTGCGCTCGGCCAGGGGCAAGACATCACCGCGCCGAAGGTAAAG 28058
Qy 369 atcagctcccgagcccaagcgcgtgagggcgacatgaaagcgcgtg 413
Db 28057 ATCATGCGGTGAAGGCCACCGCGGTGCGCAGACATGCGGCGCG 28013

Sat Apr 6 11:03:28 2002

us-09-621-781-1_copy_1507_1970_1.rni

Page 8

Search completed: April 5, 2002, 08:48:45
Job time: 1548 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2002, 08:22:57 ; Search time 1414.97 Seconds
(without alignments)
3523.779 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464
Sequence: 1 gaagagagaagagatggga.....ttggcaaccgaacgagac 464

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapept 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estfun:*

2: em_esthum:*

3: em_estlin:*

4: em_estom:*

5: em_estpl:*

6: em_estba:*

7: em_estro:*

8: em_estov:*

9: em_hlc:*

10: qb_estl:*

11: qb_est2:*

12: qb_hlc:*

13: qb_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_rtd:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	155	33.4	623	10	AW320574 us020f09.y
2	135.4	29.2	575	10	AA472295 vnh1c05.r
3	135.4	29.2	615	11	W62865 md85f08.r1
4	134	28.9	542	10	AW762360 us56h03.y
5	121.6	26.2	573	10	AA104251 mo50h03.r
6	114.6	24.7	461	10	AA920479 vy53c08.r
7	111.8	24.1	908	11	BF182426
8	108	23.3	708	10	BE573350
9	107.8	23.2	596	10	BE573255
10	103.8	22.4	446	10	AA684447
11	96	20.7	1136	10	BE573733
12	86.4	18.6	574	11	BE100299

13	84	18.1	778	11	BG535582
14	82.8	17.8	773	11	BF106314
15	81.2	17.5	642	11	AU126992
16	81.2	17.5	714	11	BG621191
17	81.2	17.5	714	11	BF311344
18	81.2	17.5	719	10	AU137751
19	81.2	17.5	747	2	BG752598
20	81.2	17.5	882	11	B1260970
21	80.2	17.3	411	10	AU136946
22	80.2	17.3	864	11	BF238180
23	80.2	17.3	558	11	BF478190
24	80	17.2	721	10	AL134765
25	79.6	17.2	516	10	AU128763
26	79.6	17.2	901	10	AL554828
27	78	16.8	383	10	BE246464
28	75.8	16.3	836	10	BE317726
29	75.4	16.3	186	10	AA381292
30	75.4	16.3	1035	10	BE613683
31	75.4	16.3	1068	11	BF037888
32	74.4	16.0	927	10	AL543823
33	72.2	15.6	752	11	B1148394
34	71.4	15.4	613	11	BG476937
35	70.2	15.1	749	11	BG500419
36	69.4	15.0	430	11	R93673
37	68.4	14.7	872	10	BE253631
38	67.2	14.5	300	10	AU098820
39	67.2	14.5	306	10	AU076649
40	66.8	14.4	695	11	B124257
41	65.8	14.2	936	11	BF128750
42	65.4	14.1	543	11	B1113465
43	64.4	13.9	476	11	BG394934
44	63.6	13.7	592	10	BE313235
45	59.2	12.8	1261	11	BG425501

ALIGNMENTS

RESULT 1
us020f09.y1 NCI-CGAP Mam6 Mus musculus cDNA clone IMAGE:2609705 5' similar to qb:x51829 Mouse myeloid differentiation primary response mRNA encoding (MOUSE);, mRNA sequence.

ACCESSION
AM320574
VERSION
AM320574.1 GI:6750118

KEYWORDS
EST.

SOURCE
house mouse.

ORGANISM
Mus musculus

REFERENCE
AUTHORS
TITLE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT
Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbp/image/image.html

Seq primer: -40RP from Gibco

High quality sequence stop: 434.

Location/Qualifiers

FEATURES
source
1..623
/organism="Mus musculus"
/strain="FVB/N"

```
/db_xref="taxon:10090"
/clone="IMAGE:2609705"
/clone_id="NCI CGAP Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/organism="mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by life technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT      137 a      197 c      173 g      115 t      1 others
ORIGIN

Query Match
Best Local Similarity 33.4%; Score 155; DB 10; Length 623;
Matches 181; Conservative 0; Mismatches 20; Indels 4; Gaps 1;

Y 262 tgttgccagagattgactcagcttcgacgtgtggaagattacatgcgagcccgcg 321
|||||
Db 4 tgttgccagagattgactcagcttcgacgtgtggaagattacatgcgagcccgcg 63
|||||

QY 322 actccgacccctt-----gcgggacagccttgcgacagcccgtagacatcagctcc 377
|||||
Db 64 ACCCGCATCCCTTGGCCGCGGACGACGCTTGTACAGCCTGTGAACATTGGCTCC 123
|||||

QY 378 ccgagcccaagcgtgagcgagacgaacgcgctgagccttgcgacatccgagcccg 437
|||||
Db 124 CCGAGCCCGACGCGCTAGGGCGACATGACCCCGCTTGGCGAGCAGTCCGAGCCAC 183
|||||

QY 438 gacgcctttgccaacgcgacccg 462
|||||
Db 184 GATCGCTTTTGGCACCAACCGG 208
|||||

RESULT 2
AA472295 575 bp mRNA EST 18-JUN-1997
LOCUS AA472295
DEFINITION Wh01c05.r1 Soares mammary_gland_NBMG Mus musculus cDNA clone
IMAGE:874184 5' similar to gb:X51829 Mouse myeloid differentiation
primary response mRNA encoding (MOUSE);, mRNA sequence.
ACCESSION AA472295
VERSION AA472295.1 GI:2200286
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 575)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HM1 Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HM1 Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:513664
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 377.
location/qualifiers
1..575
/organism="Mus musculus"
/strain="C57Bl/6J"

/db_xref="taxon:10090"
/clone="IMAGE:874184"
/clone_id="Soares_mammary_gland_NBMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/organism="mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - Oligo(dT)
primer (5'
TGTACCAATCTGAAGTGGAGCGCGCGAAGCTTTTGTGTGTGTGTGTGTGTGT
T 3'); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT      127 a      183 c      161 g      104 t
ORIGIN

Query Match
Best Local Similarity 29.2%; Score 135.4; DB 10; Length 575;
Matches 159; Conservative 0; Mismatches 16; Indels 4; Gaps 1;

Y 288 agctgtggaagattactgcgagaccccgagcttcgcattccctt-----gcgggga 343
|||||
Db 4 AGTTGTGGAAGATTACTGCGATATCCCGCGACCCCGCATCCCTTGGCGCGCGGA 63
|||||

QY 344 cagccttgcgacagcccgtagacatcagctcccgagcccgctgagggcgacat 403
|||||
Db 64 CAGCTTTGCTACACCGCTGTGAACATTGCTCCCGGAGCCCGCTGAGGGCGACAT 123
|||||

QY 404 gaacgcgtgccttgtagagacatccgagccacagatcgtcttggcaaacgcgacccg 462
|||||
Db 124 GAACCGCTGCGCTTGCAGCAGTCCGAGCCACGATGCTTTTGGCAACACGACCGG 182
|||||

RESULT 3
W62865 615 bp mRNA EST 07-JUN-1996
LOCUS W62865
DEFINITION md85f08.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:375207 5' similar to gb:X51829 Mouse myeloid
differentiation primary response mRNA encoding (MOUSE);, mRNA
sequence.
ACCESSION W62865
VERSION W62865.1 GI:1369614
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 615)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HM1 Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HM1 Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:236639
Seq primer: ETPPrimer
High quality sequence stop: 344.
```

FEATURES

SOURCE

Location/Qualifiers

1. .615

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:375207"

/clone_1lb="Soares mouse embryo NDM13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I. Site_2: Eco RI. 1st strand cDNA was primed with a Not I. oligo(dT) primer 15' TGTTCACATCTGAGTGGAGCGCGCGGAATTTTCTTTTCTTTT T 3', on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne State Univ., from 2 1) double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 145 a 130 c 165 g 115 t

ORIGIN

Query Match 29.2%; Score 135.4; DB 11; Length 615;
Best Local Similarity 88.8%; Pred. No. 1.3e-28;
Matches 159; Conservative 0; Mismatches 16; Indels 4; Gaps 1;

Oy 288 agcttgtagaattacatgagagaccccgagacccgacccctt---gcgggga 343
|||||
Db 3 AGTTGTGGAAGATACATCGATATCCCGCGACCCCGCATCCCTTGGCGGCGGA 62
|||||
Oy 344 cagccttgagacagcccgtagacatcagtcgccgagcccgcccgctgagggagacat 403
|||||
Db 63 CAGCTTGTGTAAGACCTGTGAACATTCGTCGCCGAGCCGACGCTGAGGGCGACAT 122
|||||
Oy 404 gaaagcgtgagccttgagagcaatccgagcccgagatcgcttttggcaaacgagacgg 462
|||||
Db 123 GAAACCGCTGGCTTCGCGAGAGTCGAGCCAGCATCGCTTTGGCAACAGAACCGG 181
|||||

RESULT 4

LOCUS

DEFINITION

AM762360 542 bp mRNA EST 04-MAY-2000
ur56h03.y1 NCI-CGAP_Mam3 Mus musculus cDNA clone IMAGE:3154325 5'
similar to gb:X51829 Mouse myeloid differentiation primary response
mRNA encoding (MOUSE);, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 542)
NCI-CGAP <http://www.nci.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other ESTs: ur56h03.x1
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
image.llnl.gov/image/html/tresources.shtml

FEATURES

SOURCE

MGI:1057081
Seq primer: -40RP from Gibco
High quality sequence stop: 374.
Location/Qualifiers

1. .542

/organism="Mus musculus"

/strain="129 - C57/B6 - FVB"

/db_xref="taxon:10090"

/clone="IMAGE:3154325"

/clone_1lb="NCI-CGAP_Mam3"

/tissue_type="tumor, gross tissue"

/dev_stage="10 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI. Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."

BASE COUNT 114 a 171 c 159 g 98 t

ORIGIN

Query Match 28.9%; Score 134; DB 10; Length 542;
Best Local Similarity 89.2%; Pred. No. 3.1e-28;
Matches 157; Conservative 0; Mismatches 15; Indels 4; Gaps 1;

Oy 291 ttgtgagaattacatgagagaccccgagacccgacccctt---gcgggagacg 346
|||||
Db 1 TTGTGGAAGATACATCGATATCCCGCGACCCCGCATCCCTTGGCGGCGGACAG 60
|||||
Oy 347 cctttgagacagcccgtagacatcagtcgccgagcccgcccgctgagggagacatgaa 406
|||||
Db 61 CTTTGTGTAAGACCTGTGAACATTCGTCGCCGAGCCGACGCTGAGGGCGACATGAA 120
|||||
Oy 407 cgcgtgagccttgagagcaatccgagcccgagatcgcttttggcaaacgagacgg 462
|||||
Db 121 CCCGCTGCTTCGCGAGACGTCGAGCCAGCATCGCTTTGGCAACAGAACCGG 176
|||||

RESULT 5

LOCUS

AA104251 573 bp mRNA EST 29-OCT-1996
mo50b03.r1 Life Tech mouse embryo 10 5dpc 10665016 Mus musculus
cDNA clone IMAGE:556973 5' similar to gb:X51829 Mouse myeloid
differentiation primary response mRNA encoding (MOUSE);, mRNA
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 573)
Marrin, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubague, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Weising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:337765
Seq primer: -28M13 rev1 from Amersham

416 cttagagcaatccgacccacgatacgtcttcttgcaaacggaacgg 462

LOCUS	574 bp	EST	29-JAN-2001
DEFINITION	BG100299	mRNA	
DEFINITION	u84f03.y1	McCarrey Eddy type B spermatogonia Mus musculus cDNA	
DEFINITION	clone IMAGE:3655420	5' similar to SW-016.MOUSE p15564 MELOID	
DEFINITION	DIFFERENTIATION PRIMARY RESPONSE PROTEIN MYD116.	.; mRNA sequence	

LOCUS	6G355582	778 bp	mRNA	EST	03-APR-2001
DEFINITION	602563276F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4687939 5', mRNA sequence.				
ACCESSION	BG355582				
VERSION	BG35582.1 GI:13527127				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 778)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. CDNA Library Preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution by: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L10M1501 row: h column: 20 High quality sequence stop: 510. Location/Qualifiers 1..778 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4687939" /clone_1ID="NIH_MGC_77" /lab_host="DH10B (T1 phage-resistant)" /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1: 3'f (ggcgcctcgcc); Site_2: 5'f (ggccatttgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCATATGAGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCCGAGATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."				
BASE COUNT	166 a 239 c 249 g 124 t				
ORIGIN					
Query Match	18.1%; Score 84; DB 11; Length 778;				
Best Local Similarity	69.4%; Pred. No. 9.8e-14;				
Matches 129; Conservative	0; Mismatches 55; Indels 2; Gaps 1,				
OY	279	tcagttgcagcttcttggaagattcatcagcagacccgcgagctcgcaccccttgc	338		
DB	9	TCGGTTCATCCATCCAGTGTGATCTTATCCAGACGCTGCACGCCGCCCTTGT	68		
OY	339	cggagacgcttgcagacacccgctggagacatcacgt--ccccgagcccaagcctgag	396		
DB	69	CCCCACGGACTTGAAGCGACGCCGAGATATCTGTAGTTACTGTGAG6CCGACGCCCTGAGG	128		
OY	397	gcgacatgaagcgcctgagccttggagagcaatccgagaccagatcgttttggacaacg	456		
DB	129	GTTGAATGTAACCGCTGGCTCTCCATAACCGTCCGACCTGTGATGCGCTTCTGCAACCG	188		
OY	457	aacgg 462			
DB	189	AACCG 194			
RESULT 14					
LOCUS	BF106314 773 bp mRNA EST 19-OCT-2000				
DEFINITION	601823528F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4043190 5', mRNA sequence.				
ACCESSION	BF106314				

Source

Query/Match	17.8%	Score	82.8	DB	11	Length	773
Best Local Similarity	71.5%	Pred. No.	2.2e-13				
Matches	123	Conservative	0	Mismatches	47	Indels	2
						Gaps	1

RESULT	15				
LOCUS	AUI26992				
DEFINITION	AUI26992	642 bp	mRNA	EST	23-OCT-2000
ACCESSION	AUI26992	NT2RP2	Homo sapiens	cdna clone	NT2RP2000346 5', mRNA
VERSION	AUI26992				
KEYWORDS	AUI26992.1	GI:10951708			
SOURCE	human.				
ORGANISM	Homo sapiens				

REFERENCE
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
1 (bases 1 to 642)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Euzoa: Euarchonta: Platyrrhini: Cebidae: Leontideini: Aotus.
Euzoa: Euarchonta: Platyrrhini: Cebidae: Leontideini: Aotus.
Euzoa: Euarchonta: Platyrrhini: Cebidae: Leontideini: Aotus.

Source

Query Match	17.5%;	Score 81.2;	DB 10;	Length 642;
Best Local Similarity	70.9%;	Pred. No. 6e-13;		
Matches 122;	Conservative	0;	Mismatches 48;	Indels 2;
				Gaps 1;

QY	293	gltgaagattacalgcgagaccocgcgcgacatccgacatcccttgcgcggagacagcctttg	352
Db	23	gltgttgcattctttatgcgaagcgcgcgcgaccccgccgcgcttgatgcgcacgacacttg	82
QY	353	cgaagaccgttggaactcaacgtt--cccgaagcccacgccttgatggcgaatgaagaagcg	410
Db	83	agcgacgcggagatccttgatggtactcggagccgcgcgcgcttagaggtgagatgaacgcg	142
QY	411	ctgcgccttgagaacatcgcgaccacgaatcgcctttggcgaacgaacgaagcg	462
Db	143	ctggccctcccttaacgcgcggacctgtgatgcgcttcttgcgagacacgaacgg	194

Search completed: April 5, 2002, 08:46:52
Job time: 1435 sec

C	12	37	8.0	1606	21	AAC41089	Arabidopsis thaliana
C	13	36.2	7.8	494	20	AAV66970	EST clone B188.
C	14	33.8	7.3	28804	17	AACT3329	Sphingon biosynthe
C	15	33.8	7.3	28804	18	AAAT9474	Sphingomonas genus
C	16	33.8	7.3	28804	20	AAAV9812	Sphingomonas 588 s
C	17	33.8	7.3	28804	20	AAV81474	Chromosomal fragme
C	18	33.2	7.2	5160	21	AAAF21033	Human low adenosin
C	19	33.2	7.2	13548	21	AAA34911	Human adenosine re
C	20	33.2	7.2	13548	21	AAAF21037	Human low adenosin
C	21	33.2	7.2	13548	21	AAA34915	Human adenosine re
C	22	32.4	7.0	1598	16	AAVT04953	Mouse Kvl.7 voltae
C	23	31.8	6.9	3747	22	AAFT26918	Streptococcus muta
C	24	31.6	6.8	3496	22	AAHA6183	Phosphoryl transpo
C	25	31.6	6.8	6045	22	AAHI8244	Human cDNA sequenc
C	26	31.4	6.8	474	22	AAI39048	Probe #7734 used t
C	27	31.4	6.8	2729	21	AAA38443	Murine desmin gene
C	28	31.4	6.8	2730	21	AAZ93780	Mouse desmin promo
C	29	31.4	6.8	109519	22	AAAS08693	Micromonospora DNA
C	30	31.2	6.7	1044	21	AAAS1697	Burkholderia cepac
C	31	31.2	6.7	1044	21	AAZ51702	Burkholderia multi
C	32	31.2	6.7	1044	21	AAZ51701	Burkholderia multi
C	33	31.2	6.7	1044	21	AAZ51702	Burkholderia multi
C	34	31.2	6.7	26664	21	AAAG6207	Human prostate can
C	35	31	6.7	905	22	AAH74539	Nucleotide sequenc
C	36	31	6.7	905	22	AAH74541	Nucleotide sequenc
C	37	31	6.7	1294	20	AAZ06824	Nucleotides albid
C	38	31	6.7	1294	22	AAH74537	Nucleotide sequenc
C	39	30.8	6.6	1718	18	AAH93529	Maturase and Coat
C	40	30.8	6.6	1718	19	AAV18845	PAR-1 containing b
C	41	30.8	6.6	1718	20	AAV18847	Bacteriophage ms2
C	42	30.8	6.6	1867	18	AAAT93530	PAR-2 as "Armoured
C	43	30.8	6.6	1867	19	AAV18848	PAR-2 containing b
C	44	30.8	6.6	1867	20	AAV78450	HIV gag gene (QSKR
C	45	30.8	6.6	106746	21	AAA10225	Human PCTA-1 genom

ALIGNMENTS

RESULT
1

AAV65766 standard; cDNA; 2614 BP.
AAV65766;

XX AC AAV65766;
XX DT 02-FEB-1999 (first entry)
XX DE Rat progression elevated gene-2 (PEG-3) promoter region.
XX KM Progression elevated gene-3; PEG-3; rat; tumour progression;
KM DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
KM brain cancer; cervix cancer; prostate cancer; lung cancer;
KM colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy;
KW promoter; ss.
XX OS Rattus sp.
XX XX
XX Key Location/Qualifiers
FH misc_feature /tag= a
FT /label= PEA3
FT 2128..2133
FT misc_feature /tag= b
FT /label= PEA3
FT 2404..2409
FT misc_feature /tag= c
FT /label= PEA3
FT 338..344
FT misc_feature /tag= d
FT /label= EZA
FT 351..357
FT misc_feature /tag= e
FT /label= GRE

XX Novel vectors useful for studying the progression of cancer -
 PT Disclosure; Fig 14 A-B; 251pp; English.
 XX
 XX
 CC The invention relates to an inducible progression-elevated gene-3 (PEG-3
 CC gene) regulatory region functionally linked to a gene encoding a product
 CC that causes or may be induced to cause the death or inhibition of cancer
 CC cell growth. A vector of the invention which contains a gene encoding
 CC thymidine kinase or a product which causes the cell to express a
 CC specific antigen can be administered along with gancyclovir or acyclovir,
 CC or an antibody or fragment to the antigen, respectively, to treat cancer
 CC in a subject. The PEG-3 gene is useful for generating new cloning and
 CC expression vectors, transfected cells, and for developing methods for
 CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as
 CC a source of primers and probes to study the progression of cancer, and to
 CC detect the presence of the gene. The present sequence represents the
 CC nucleotide sequence of the rat PEG-3 gene promoter.
 XX
 SQ Sequence 2614 BP; 737 A; 615 C; 706 G; 556 T; 0 other;

Query Match 92.2%; Score 428; DB 20; Length 2614;
 Best Local Similarity 99.4%; Pred. No. 1.8e-128;
 Matches 461; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 1 gaaagagaagaagaatggagacagcatgtgactcctcgatgaagtgtgctgtgtctaa 60
 DB 1512 gaaagagaagaagaatggagacagcatgtgactcctcgatgaagtgtgctgtctaa 1571
 QY 61 aagtttcgagattgaagcgtctctctgatttgagccaagagacgccttggaagccacg 120
 DB 1572 aagtttcgagattgaagcgtctctctgatttgagccaagagacgccttggaagccacg 1631
 QY 121 gtgacctcaagagcccggaatctccgcgagaaattcagtggttttccctctccacc 180
 DB 1632 gtgacctcaagagcccggaatctccgcgagaaattcagtggttttccctctccacc 1691
 QY 181 ttttcagggaattcgaaactcgcgtctccggtgacgcagcatagagcgtgctgca 240
 DB 1692 ttttcagggaattcgaaactcgcgtctccggtgacgcagcatagagcgtgctgca 1749
 QY 241 ctataaactccgggtgtagctgtgttgcgagatttgactcagttcgagcttctggaaga 300
 DB 1750 ctataaactccgggtgtagctgtgttgcgagatttgactcagttcgagcttctggaaga 1809
 QY 301 ttacatgagagaccgcgcgcgactccgactcctttgctggagacagcctttgcagagcc 360
 DB 1810 ttacatgagagaccgcgcgcgactccgactcctttgctggagacagcctttgcagagcc 1869
 QY 361 cgtgagacatcagttcccccagagcccccgcgtgagggcgagacagcgctgagccttga 420
 DB 1870 cgtgagacatcagttcccccagagcccccgcgtgagggcgagacagcgctgagccttga 1928
 QY 421 gagcaatccgagaccacagatcgcttttgcaaacgcgagcagac 464
 DB 1929 gagcaatccgagaccacagatcgcttttgcaaacgcgagcagac 1972

RESULT 3
 AAV65765
 ID AAV65765 standard; cDNA; 2111 BP.
 XX
 AC AAV65765;
 XX
 DT 02-FEB-1999 (first entry)
 XX
 DE Human progression elevated gene-2 (PEG-3) cDNA.
 XX
 KW Progression elevated gene-3; PEG-3; human; tumour progression;
 KW DNA repair; tumorigenesis; angiogenesis; metastasis; melanoma;
 KW brain cancer; cervix cancer; prostate cancer; lung cancer;
 KW colorectal cancer; neuroblastoma; glioblastoma; diagnosis; therapy;

KW SS.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 294..2030
 FT /*tag= a
 FT
 PN W09842315-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 20-MAR-1998; 98WO-US05793.
 XX
 PR 21-MAR-1997; 97US-0821818.
 XX
 PA (GENO-) GENQUEST INC.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Fisher PB, Goldstein NI, Su Z, Zhang N;
 XX
 DR WPI. 1998-557025/47.
 XX P-PSDB; AAW79938.
 XX
 PT New isolated Progression Elevated Gene-3 - used to develop products
 PT for e.g. modulating DNA damage and repair pathways, cancer
 PT progression or oncogene mediated transformation and angiogenesis.
 XX
 XX Claim 4; Fig 13A-C; 225pp; English.
 XX
 CC This nucleotide sequence includes an open reading frame encoding
 CC human progression elevated gene-3 (PEG-3) protein (see AAW79937).
 CC PEG-3 clones were isolated from an MCF-7 cDNA library using rat
 CC PEG-3 cDNA (see AAV65764) as probe. 5'RACE (see AAV65767-68) was used
 CC to generate full-length cDNA. PEG-3 mRNA is highly expressed in
 CC most human tumour cell lines. Probes derived from the gene can be
 CC used to monitor tumour progression. Antisense sequences can be
 CC used to inhibit tumour progression. Cells in which a reporter gene
 CC is under control of the promoter (see AAV65766) of the rat PEG-3 gene
 CC can be used in methods for identifying agents that modulate PEG-3
 CC expression or the ability of PEG-3 to induce progression, or for
 CC determining whether an agent is capable of inhibiting DNA damage
 CC and repair pathways, cancer progression or oncogene-mediated
 CC transformation. Compounds that induce DNA damage or which regulate
 CC angiogenesis can also be identified using such cells. Transgenic
 CC animals and vaccines comprising PEG-3 polypeptides and an immune
 CC response enhancer are also claimed. The methods can be applied to
 CC a progression phenotype comprising anchorage independent growth,
 CC tumorigenesis, angiogenesis or metastasis, to melanoma, brain,
 CC cervical, prostate, lung or colorectal cancer, neuroblastoma or
 CC glioblastoma.
 XX
 SQ Sequence 2111 BP; 533 A; 607 C; 575 G; 396 T; 0 other;

Query Match 28.3%; Score 131.4; DB 19; Length 2111;
 Best Local Similarity 99.2%; Pred. No. 1.8e-32;
 Matches 132; Conservative 1; Indels 0; Gaps 0;

QY 332 cctttgcggagacagccttttgagacagccgtgagacatcagttcccgagccagccg 391
 DB 120 cctttgcggagacagccttttgagacagccgtgagacatcagttcccgagccagccg 179
 QY 392 ttagggcgacatgaacgcgctgcttgagagaaatccggaccacagcatcgtttgca 451
 DB 180 ttagggcgacatgaacgcgctgcttgagagaaatccggaccacagcatcgtttgca 239
 QY 452 aaccgagaccgagac 464
 DB 240 aaccgagaccgagac 252

RESULT 4

```

AA23029
ID AA23029 standard; cDNA; 2111 BP.
XX
AC AA23029;
XX
DT 17-JAN-2000 (first entry)
XX
DE Human progression elevated gene-3 (PEG-3) cDNA sequence.
XX
KM Progression-elevated gene-3; PEG-3 gene; regulatory region; cell death;
KW cell growth; cancer; thymidine kinase; gancyclovir; acyclovir; human; ss.
XX
OS Homo sapiens.
XX
PN W09949898-A1.
XX
PD 07-OCT-1999.
XX
PF 31-MAR-1999; 99WO-US07199.
XX
PR 31-MAR-1998; 98US-0052753.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB;
XX
PI WPI; 1999-591184/50.
XX
DR P-PSDB; AAY41104.
XX
PT Novel vectors useful for studying the progression of cancer -
XX
PS Disclosure; Fig 13A-C; 251pp; English.
XX
CC The invention relates to an inducible progression-elevated gene-3 (PEG-3
CC gene) regulatory region functionally linked to a gene encoding a product
CC that causes or may be induced to cause the death or inhibition of cancer
CC cell growth. A vector of the invention which contains a gene encoding
CC thymidine kinase or a product which causes the cell to express a
CC specific antigen can be administered along with gancyclovir or acyclovir,
CC or an antibody or fragment to the antigen, respectively, to treat cancer
CC in a subject. The PEG-3 gene is useful for generating new cloning and
CC expression vectors, transfected cells, and for developing methods for
CC cultured growth of such cells. The PEG-3 polynucleotide is also useful as
CC a source of primers and probes to study the progression of cancer, and to
CC detect the presence of the gene. The present sequence represents the
CC nucleotide sequence of the human PEG-3 cDNA.
XX
SQ Sequence 2111 BP; 533 A; 608 C; 575 G; 395 T; 0 other;

Query Match 28.3%; Score 131.4; DB 20; Length 2111;
Best Local Similarity 99.2%; Pred.No.1.8e-32;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 332 ccttgcggagacagccttgcagacagccgtgagacatcagtcctccgagcccaagcc 391
DB 120 ccttgcggagacagccttgcagacagccgtgagacatcagtcctccgagcccaagcc 179
QY 392 tgagggcgacatgaacgcgtgccttgagagcaatccggaccacagatcgctttggca 451
DB 180 tgagggcgacatgaacgcgtgccttgagagcaatccggaccacagatcgctttggca 239
QY 452 aacggaaccggac 464
DB 240 aacggaaccggac 252

RESULT 5
ID AAX87383 standard; cDNA; 2111 BP.
XX
AC AAX87383;
XX

```

```

DT 08-OCT-1999 (first entry)
XX
DE Human progression elevated gene-3 (PEG-3) cDNA.
XX
KM Progression elevated gene-3; PEG-3; human; angiogenesis; cancer;
KW therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 294..2030
FT /tag= a
XX
PN W09937776-A1.
XX
PD 29-JUL-1999.
XX
PF 26-JAN-1999; 99WO-US01623.
XX
PR 29-JAN-1998; 98US-0072941.
XX
PR 26-JAN-1998; 98US-0072469.
XX
PA (GENO-) GENQUEST INC.
XX
PI Fisher PB, Zhang N;
XX
PI WPI; 1999-458694/38.
XX
DR P-PSDB; AAY06514.
XX
PT Modulation of angiogenesis by altering the expression and/or
PT activity of a progression-associated protein, especially for cancer
PT treatment
XX
PS Claim 1; Page 71-74; 81pp; English.
XX
CC This is the nucleotide sequence of a cDNA clone corresponding to
CC the human progression elevated gene-3 (PEG-3). It codes for a
CC 578-amino acid protein (see AAY06514), and represents a new member
CC of the gadd45/Wyd116 gene family. The human PEG-3 cDNA was
CC isolated from a MCF-7 cDNA library using rat PEG-3 (see AAX87387) as
CC probe. Human PEG-3 is generally expressed in cells that are in
CC progression, including most human tumour cell lines. A claimed
CC method for modulating angiogenesis in an organism comprises
CC administering an agent that alters expression and/or activity of a
CC PEG-3 protein, such as an antisense polynucleotide or antibody.
CC Also claimed is a method for determining whether an agent
CC modulates angiogenesis, where the candidate agent is present
CC within a combinatorial small molecule library. Vaccines and
CC pharmaceutical compositions comprising such compounds are also
CC provided and may be used to prevent angiogenesis, especially
CC related to cancer cell progression.
XX
SQ Sequence 2111 BP; 533 A; 608 C; 575 G; 395 T; 0 other;

Query Match 28.3%; Score 131.4; DB 20; Length 2111;
Best Local Similarity 99.2%; Pred.No.1.8e-32;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 332 ccttgcggagacagccttgcagacagccgtgagacatcagtcctccgagcccaagcc 391
DB 120 ccttgcggagacagccttgcagacagccgtgagacatcagtcctccgagcccaagcc 179
QY 392 tgagggcgacatgaacgcgtgccttgagagcaatccggaccacagatcgctttggca 451
DB 180 tgagggcgacatgaacgcgtgccttgagagcaatccggaccacagatcgctttggca 239
QY 452 aacggaaccggac 464
DB 240 aacggaaccggac 252

RESULT 6

```

[illegible]

xx	RESULT 7
xx	ID AAH05790
xx	AAH05790 standard; cDNA: 642 BP.
xx	AC AAH05790;
xx	DT 26-JUN-2001 (first entry)
xx	DE Human cDNA clone (5'-primer) SEQ ID NO:2625.
xx	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
xx	XX Homo sapiens.
xx	PN EPI074617-A2.
xx	PD 07-FEB-2001.
xx	PF 28-JUL-2000; 2000EP-0116126.
xx	PR 29-JUL-1999; 99JP-0248036.
xx	PR 27-AUG-1999; 99JP-0300253.
xx	PR 11-JAN-2000; 2000JP-0118776.
xx	PR 02-MAY-2000; 2000JP-0185767.
xx	PR 09-JUN-2000; 2000JP-0241899.
xx	PA (HELI-) HELIX RES INST.
xx	PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
xx	PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
xx	DR WPI; 2001-318749/34.
xx	PT primer sets for synthesizing polynucleotides, particularly the 5602
xx	PT full-length cDNAs defined in the specification, and for the detection
xx	PT and/or diagnosis of the abnormality of the proteins encoded by the
xx	PT full-length cDNAs -
xx	PS Claim 1; SEQ ID 2625; 2537bp + CD ROM; English.
xx	CC The present invention describes primer sets for synthesizing 5602
xx	CC full-length cDNAs defined in the specification. Where a primer set
xx	CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
xx	CC to the complementary strand of a polynucleotide which comprises one of
xx	CC the 5602 nucleotide sequences defined in the specification, where the
xx	CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
xx	CC of an oligonucleotide comprising a sequence complementary to the
xx	CC complementary strand of a polynucleotide which comprises a 5'-end
xx	CC sequence and an oligonucleotide comprising a sequence complementary to a
xx	CC polynucleotide which comprises a 3'-end sequence, where the
xx	CC oligonucleotide comprises at least 15 nucleotides and the combination of
xx	CC the 5'-end sequence/3'-end sequence is selected from those defined in
xx	CC the specification. The primer sets can be used in antisense therapy and
xx	CC in gene therapy. The primers are useful for synthesizing polynucleotides,
xx	CC particularly full-length cDNAs. The primers are also useful for the
xx	CC detection and/or diagnosis of the abnormality of the proteins encoded by
xx	CC the full-length cDNAs. The primers allow obtaining of the full-length
xx	CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
xx	CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
xx	CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
xx	CC represent oligonucleotides, all of which are used in the exemplification
xx	CC of the present invention.
xx	SQ Sequence 642 BP; 131 A; 200 C; 190 G; 115 T; 6 other;
QY	Query Match 17.5%; Score 81.2; DB 22; Length 642;
DB	Best Local Similarity 70.9%; Pred. No. 2e-16;
	Matches 122; Conservative 0; Mismatches 48; Indels 2; Gaps 1
	293 gttgaagataactatgagagcccgagcgagctcgcgactcccttcgggagacgcttgg 352
	111
	23 gttgtatctcttcgaagagctgcaagcccgagcccgcttcgttcgccaagcagcttg 82

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142350.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151330.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158322.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.


```

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

```

Query Match 8 0%; Score 37; DB 21; Length 1606;
 Best Local Similarity 57.3%; Pred. No. 0.059; 50; Indels 0; Gaps 0;
 Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```

QY 37 gatgaattggcgtgcttgcctcaaaagttctggaattgacggcctctgtgattgac 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 267 GAAGACGTTGGCGATCGTGTCCACAGACGTACACAAATTGACTTCTTGACCTGTCAAA 208
QY 97 caaggaacacgcctgggaagccacgagtgacccacaaagccggaatccggaagaa 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 AAGGTACAGGCGCTTGCAAGCGCCATTAGCTTCACTGCTTGCGGGAATTGAGAGAAA 151

```

RESULT 13
 AAV86970
 ID AAV86970 standard; cDNA; 494 BP.
 XX
 AC AAV86970;
 XX
 DT 27-APR-1999 (first entry)
 XX
 DE EST clone B188.
 XX

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytic;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX MO9845435-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US06954.

XX 10-APR-1997; 97US-0835913.

XX (GENE) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racine LA, Spaulding V, Treacy M;

XX WPI, 1999-070076/06.

PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries

PS Claim 1; Page 409; 633pp; English.

CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokine activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC therapy. The EST sequences are also stated to be useful for gene
 CC therapy.

XX Sequence 494 BP; 158 A; 88 C; 91 G; 157 T; 0 other;

Query Match 7.8%; Score 36.2; DB 20; Length 494;
 Best Local Similarity 48.3%; Pred. No. 0.066;
 Matches 101; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

```

QY 20 acaacatgtacgtccgtgaagatggcgtgctgctcaaaagttctggaattgacg 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 agattataagacatcccttaagaagtttaaatgtctctgaacagacatacaggtctagt 159

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2002, 08:22:57 ; Search time 1472.67 Seconds
(without alignments)
5197.832 Million cell updates/sec

Title: US-09-621-781-1_COPY_1507_1970

Perfect score: 464
Sequence: 1 gaaagagaaagaaatggga.....tttgcaaacgacgacgac 464

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : GenBank:*

1: gb_ba:*

2: gb_hlg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_om:*

20: em_or:*

21: em_ov:*

22: em_pat:*

23: em_ph:*

24: em_pl:*

25: em_ro:*

26: em_sts:*

27: em_sy:*

28: em_un:*

29: em_vl:*

30: em_htgo_hum:*

31: em_htgo_inv:*

32: em_htgo_rnd:*

33: em_htg_hum:*

34: em_htg_inv:*

35: em_htg_rnd:*

36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	464	100.0	1940	10 AF351130	AF351130 Rattus no
2	300.6	64.8	215734	2 AC073828	AC073828 Mus muscu
3	154	33.2	216	10 MM083984	MM083984 Mus muscu
4	136.6	29.4	2275	10 MMBPBRMR	M51829 Mouse myelo
5	116.4	25.1	179538	2 AC024740	AC024740 Homo sapi
6	114.8	24.7	110000	2 AC008749_3	Continuation (4 of
7	114.8	24.7	231450	2 AC026803	AC026803 Homo sapi
8	114.2	24.6	2088	10 CRUGAD34X	L28147 Hamster Gad
9	106.2	22.9	165166	2 AC064877	AC064877 Homo sapi
10	96.2	20.7	224	10 CGU83983	U83983 Homo sapien
11	81.2	17.5	2331	6 HSB02091	HSB02091 Sequence
12	81.2	17.5	2344	6 AK001361	AK001361 Homo sapi
13	81.2	17.5	2349	6 BC003067	BC003067 Homo sapi
14	81.2	17.5	2358	9 HSB03982	H83982 Homo sapien
15	45.4	9.8	233	9 AC027139	AC027139 Homo sapi
16	37.8	8.1	151498	9 AC021822	AC021822 Homo sapi
17	37.4	8.1	202475	2 AC009973	AC009973 Homo sapi
18	37.4	8.0	91566	8 F25C20	F25C20 Arabidops
19	36.6	7.9	2397	9 HSB01096	HSB01096 Homo sapi
20	36.6	7.9	3993	9 HSB02091	AB020662 Homo sapi
21	36.6	7.9	4241	9 AB020662	AF204231 Homo sapi
22	36.6	7.9	4578	9 AF204231	AC021822 Homo sapi
23	36.6	7.9	174320	2 AC021822	AC025678 Homo sapi
24	36.6	7.9	193361	2 AC025678	AF164622 Homo sapi
25	36.2	7.8	4371	9 AF164622	AF297204 Bos tauri
26	35.4	7.6	617	11 AF297204	AF297204 Bos tauri
27	35.2	7.6	1218	3 AF326988	AF326988 Citellidia
28	35	7.5	155628	2 AC027810	AC027810 Homo sapi
29	35	7.5	203397	2 AC024619	AC024619 Homo sapi
30	34.8	7.5	148098	2 AL359315	AL359315 Homo sapi
31	34.8	7.5	152623	9 AC011595	AC011595 Homo sapi
32	34.8	7.5	164388	2 AL591647	AL591647 Homo sapi
33	34.8	7.5	173829	2 AL451077	AL451077 Homo sapi
34	34.8	7.5	181427	2 AC027209	AC027209 Homo sapi
35	34.8	7.5	183763	2 AC026900	AC026900 Homo sapi
36	34.8	7.5	192458	2 AL592284	AL592284 Homo sapi
37	34.8	7.5	219461	2 AC087864	AC087864 Homo sapi
38	34.8	7.5	227243	2 AC008569	AC008569 Homo sapi
39	34.6	7.5	6426	1 RCBUDGNS	Z46611 R. capsulatu
40	34.6	7.5	113606	8 AC087192	AC087192 Oryza sat
41	34.6	7.5	189370	1 AF010496	AF010496 Rhodobact
42	34.4	7.4	15019	1 AE006957	AE006957 Mycobacte
43	34.4	7.4	29550	1 MFW039	AL021942 Mycobacte
44	34.4	7.4	124219	9 HSL196B13	AL121968 Human DNA
45	34.4	7.4	147165	2 AC023179	AC023179 Homo sapi

ALIGNMENTS

RESULT 1

AF351130 AF351130 1940 bp DNA 09-APR-2001

LOCUS Rattus norvegicus progression elevated gene-3 protein (peg-3) gene,

DEFINITION promoter region.

ACCESSION AF351130

VERSION AF351130.1 GI:13562027

KEYWORDS

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1940)

AUTHORS Su,Z., Shi,Y. and Fisher,P.B.

TITLE Cooperation between AP1 and PEA3 sites within the progression elevated gene-3 (PEG-3) promoter regulate basal and differential expression of PEG-3 during progression of the oncogenic phenotype in transformed rat embryo cells

JOURNAL Oncogene 19 (30), 3411-3421 (2000)

MEDLINE 20374129
PUBMED 10918598
REFERENCE 2 (bases 1 to 1940)
AUTHORS Su,Z., Shi,Y., and Fisher,P.B.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2001) Urology, Columbia University, 630 West 168th Street, New York, NY 10032, USA

FEATURES
Source
1. .1940 /organism="Rattus norvegicus"
/db_xref="taxon:10116"
promoter 1. .1940 /gene="Peg-3"
1. .>1940 /gene="Peg-3"
gene /note="progression elevated gene-3"

BASE COUNT 568 a 496 c 468 g 408 t
ORIGIN

Query Match 100.0%; Score 464; DB 10; Length 1940;
Best Local Similarity 100.0%; Pred. No. 1.7e-117;
Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaaagagaagagaatggagacagatgtgactgctgtaagttggcgtgtgtctcaa 60
1477 GAAAGAGAAAGAGATGGAGACATGTGACTGCTGATGAGTTGGCGTGTGTCGAA 1536

OY 61 aagttcgcagattgcagctctctgtgattgagccaagacagccttggagaacag 120
1537 AAGTTTCGCGAGATTGACGGCTCTCTGTGATTTAGCCAGACACGCTTGGAGCCACG 1596

OY 121 gtgacctacaagcccggaatctccgcgagaatttcagttgttttctctctcacc 180
1597 GTGACCTCACAGGCCCGGGAATCTCCGCGAGAAATTCAGTGTGTTTCTCTCTCACG 1656

OY 181 ttctcaggaacttcgaaactcgcctccgcgtgtaagtcacagaagcgtgcgtcaga 240
1657 TTCTCTAGGAGCTTCGAAATCTCGGCTCTCCGGTAGCTACACATAGCGCTTCCGTGAGA 1716

OY 241 ctataactcccggtgtagctgtgtgacagattgactcaattgcagcttgtagaaga 300
1717 CTATTAACCTCCGGGTGATGTGTGGCGAGATTGACTCAGTTCCAGCTTGTGAGAGA 1776

OY 301 ttacatggagaccccgcgagctccgcatcccttgcggagacagccttggagacgc 360
1777 TTACATGCGAGACCCCGCGGACTCCGCAATCCCTTGGCGGACAGCTTTGGAGACGCG 1836

OY 361 cgtgagacatcagctcccgagcccgagcgccttggagagagacgtgagcgtgcgttga 420
1837 CGTGAGACATCAGCTCCCGAGGCCCGCAGCTTGAGGGGAGACATGAACGCGCTGCTTGA 1896

OY 421 gagcaatccggaccacagatcgcttlttgcaaacacgacggagc 464
1897 GAGCAATCCGGACCCAGATCGCTTTTGGCAAAACGACACCGGAGC 1940

RESULT 2
LOCUS AC073828 215734 bp DNA 29-JUN-2000
DEFINITION Mus musculus clone Rp23-9J18, WORKING DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION AC073828
VERSION AC073828.1 GI:8810445
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 215734)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 215734)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1740562
Center clone name: RPCI-23_9J18

Summary Statistics
Consensus quality: 196667 bases at least Q40
Consensus quality: 207971 bases at least Q30
Consensus quality: 209928 bases at least Q20
Estimated insert size: 207740; agarose-fp estimation
Estimated insert size: 213734; sum-of-ctigs estimation
Quality coverage: 6.12 in Q20 bases; agarose-fp estimation
Quality coverage: 5.94 in Q20 bases; sum-of-ctigs estimation
NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 1013: contig of 1013 bp in length
1 1014 1113: gap of unknown length
1 1114 2175: contig of 1062 bp in length
1 2176 2275: gap of unknown length
1 2276 3429: contig of 1154 bp in length
1 3430 3529: gap of unknown length
1 3530 4884: contig of 1355 bp in length
1 4885 4985: gap of unknown length
1 4985 6848: contig of 1864 bp in length
1 6849 6949: gap of unknown length
1 6949 9317: contig of 2369 bp in length
1 9318 9417: gap of unknown length
1 9418 12128: contig of 2711 bp in length
1 12129 12228: gap of unknown length
1 12229 14970: contig of 2742 bp in length
1 14971 15070: gap of unknown length
1 15071 19571: contig of 4501 bp in length
1 19572 19671: gap of unknown length
1 19672 25961: contig of 6290 bp in length
1 25962 26061: gap of unknown length
1 26062 36753: contig of 10692 bp in length
1 36754 36853: gap of unknown length
1 36854 49116: contig of 12263 bp in length
1 49117 49216: gap of unknown length
1 49217 61879: contig of 12663 bp in length
1 61880 61979: gap of unknown length
1 61980 72513: contig of 10534 bp in length
1 72514 72613: gap of unknown length
1 72613 83721: contig of 11108 bp in length
1 83722 83821: gap of unknown length
1 83822 98633: contig of 14812 bp in length
1 98634 98733: gap of unknown length
1 98734 118323: contig of 19590 bp in length
1 118324 118423: gap of unknown length
1 118424 137668: contig of 19245 bp in length
1 137669 137769: gap of unknown length
1 137769 162338: contig of 24570 bp in length
1 162339 162438: gap of unknown length
1 162439 186725: contig of 24187 bp in length
1 186726 215734: gap of unknown length
1 215734 186726: contig of 29009 bp in length.
Location/Qualifiers
1. .215734

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-9J18"
/clone_1fb="RPCI mouse BAC library 23"
BASE COUNT      54325 a 53055 c 52767 g 53583 t 2004 others
ORIGIN

```

Query Match	64.8%;	Score 300.6;	DB 2;	Length 215734;
Best Local Similarity	83.7%;	Pred. No. 1.3e-72;		
Matches 389;	Conservative 0;	Mismatches 69;	Indels 7;	Gaps 4

Oy	2	aaagagaaagaaatggygcacgcatgtgactgcctgatagaagtggggtgctgtcctcaaa	61
Db	8042	aaagaaagggaattggcaccggtgactgcccagatgaagttccgctactttgctcga	8101

Qy 62 agtctcgtgagatgcagcgctctcgtgatattgaccaaagcacgcctgggaagccacg 121
| | | | | | | | | | | | | | | | | |
Db 8102 AATTACGTGAGATCGACGGCTCTTCGA-TTGAGCCTAGGGCA CGCCTGAGAAGCTGCGG 8160

```

QY      122  tgaacctcaagaagcccggaatctctcgcgagaaattcaagtggttttcctctccacac 181
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      8161  TGACCTCACAGAGCGCGGAGTCTCCGCGAGAAATTCAAGTGTTCCTCTTTCACACT 8220

```

Qy	182 ttctcaggagacttcogaactcgcgtctccggtgacgtcagcatagcgctggctcagac 241
Db	8221 TTCCACGGACTTTCGAGACCCCGCTCTCCGGTGACGTACACACGCGCGCTCAGG 8280

Qy 212 tataaacctccggygatcgtgttcgcgcagatctgactcacgtccgcacgtcggaaagac 301
||||| || | | ||||| ||||| ||| | | ||||| ||
Db 8281 TATATAAGCGCGGTGGACGATTGTTGGCCGACATTGACTCAGCTCTGAGTTCGTGGACAGAT 8340

Qy 302 tacttgcgagaccocgcgcactccgatccctt----gccggagacgcctttcgcaca 35/
| | | | | | | | | | | | | | | | | | | | | |
Db 8341 TACATGCGATATCCGGCGCAGCCCCGCATCCCTTTGCCGGCCCGGACAGCCTTGCTACA 840U

[illegible]

Oy	418	tgagagcaatccgagaccacgatcgcttltgycaaacgcgaacg	462
Db	8460	CGGAGCAGATCCGAGCCACGATCGC-TTTGGCAACCA6AACCG	8503

RESULT 3
MM083984

DEFINITION	Mus musculus apoptosis associated protein (GADD34) gene, promoter sequence.
ACCESSION	U83394

KEYWORDS	SOURCE	ORGANISM
house mouse.		
Mus musculus		

REFERENCE
AUTHORS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
1 (bases 1 to 216)
Hollander, M.C., Zhan, Q., Bae, I. and Fornace, A.J. Jr.

JOURNAL	J. Biol. Chem.	272 (21),	13731-13737 (1997)
MEDLINE	97298078		
REFERENCE	2 (bases 1 to 216)		

TITLE Direct Submission
Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda
JOURNAL 20892, USA

```
1. . 216
source
/organism="Mus musculus"
/db_xref="taxon:10090"
```

```
/gene="GADD34"
```

```

/note="growth arrest and DNA damage-inducible gene 34"
Myd16 gene"
<1>>216
/gene="GADD34"
135, 146
/gene="GADD34"
149, 158
/misc.signal

```

```

TATA_signal . 177. .182
                /gene="GADD34"
BASE COUNT    40 a      66 c      63 g      47 t
ORIGIN

```

Query Match	33.28;	Score 154;	DB 10;	Length 216;
Best Local Similarity	85.58;	Pred. No. 7.1e-32;		
Matches 183;	Conservative 0;	Mismatches 30;	Indels 1;	Gaps 1;

Dy 68 ggcagatctgacgcgcctcctgtgaattggaccagaabacacgcctggaagcccaagcttgacct 124
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4 GTGAGATCGACGGCTCTCGA-TTGACCCTTAGGGCACCCTTGGAAGCTGCGGTGACTT 62

DY 128 cacaagcccggaatctcgcgagaatttcagtgttfttcctctccacattcca 18
||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 63 CACAGAGCCCGAGTCTCCGCGAGAAGTTCA GTGTGTTTTCCTCTTCCACTTCCA 12

Qy 188 gggacttcggaactccgcctccgcygaactcagcatgcgtcgtcagacataaa 24
 |||||
 123 gggacttccgagaccccccctctccgggtacgtacgacacaccccccgtcagggatTAA 18
 Db

Qy	248	ctcccggtgatcgtgtgcgagatlgactca	281
Db	183	AGCCGCTGACGATGTTGGCCAGATTGAGTCA	216

RESULT 4
MMMDPRMR
13-SEP-1003

DEFINITION	Mouse myeloid differentiation primary response mRNA encoding Myd11b protein.
ACCESSION	X51829

KEYWORDS	Myd16 gene; myeloid differentiation primary response gene.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 2275)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
Liebermann, D.A.

JOURNAL
Submitted (12-FEB-1990) Lieberman D.A., University of Pennsylvania
School of Medicine, Dept of Biochemistry & Biophysics, Philadelphia
PA 19104-6059, USA

AUTHORS Lord, K. A., Hoffman-Liebermann, B. and Liebermann, D. A.
TITLE Sequence of MYD16 cDNA: a novel myeloid differentiation primary response gene induced by IL6

MEDLINE	90251472	Location/Qualifiers
FEATURES		1. .2275
source		

```

/strain="SL"
/db_xref="taxon:10090"
/clone_lib="lambda Zap"

```

```

CDS      229..2202
         /note="MYD16 protein (AA 1-657)"
         /codon_start=1

```

```

/db_xref="GI:53041"
/db_xref="SWISS-PROT:P17564"

```

	/note="polyA site"			
BASE COUNT	577 a	652 c	636 g	410 t
ORIGIN				

Query Match	29.48;	Score 136.6;	DB 10;	Length 2275;
Best Local Similarity	87.68;	Pred. No. 3.6e-27;		
Matches 162;	Conservative	0;	Mismatches 19;	Indels 4;
				Gaps 1

Dy
202 gcttcgcaagctcgttgtaagaattacatcgtagaacccgcgcgaactccgcatacctt----g 337

Dd
1 GCCTTGAGTTGTGSAAGATTACATCGCATATCCGCCGACCCGGCATTCCCTTGGCGG 60

61 CCGGACAGCCTTTGCTACAGCCTGTGAACATTTGCGTCCCGAGCCCCACAGCCTGAGGG 120

Db 121 CGACATGAACCCGCTTGGCAGACAGTCCGGACCCACGATCGCTTTGGCAACCAGA 180

Db 181 ACCGG 185

RESULT	5		
AC024740/c			
LOCUS	AC024740	179538 bp	DNA
DEFINITION	Homo sapiens chromosome 19 clone RP11-801D6, WORKING DRAFT		
ACCESSION	SEQUENCE, 17 unordered pieces.		
	AC024740		
	AC024740		

VERSION	AC024/40.3	GI:85/0405
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 179538)
Waterston, R.H.

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 179538)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission

COMMENT
On Jun 17, 2000 this sequence version replaced q1:7263917.
MO 63108, USA
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
Missouri 63108, USA
Submitted (14 MAR 2000) genome sequencing center, Washington

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0801006
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid, 0%
Chemistry: Dye-primer ET, 97% of reads

```

```

Chemistry: Dye-terminator Big Dye; 3% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 173850 bases at least Q40
Consensus quality: 175557 bases at least Q30
Consensus quality: 176446 bases at least Q20
Insert size: 242000; agarose-fp
Insert size: 177938; sum-of-contigs
Quality coverage: 5.32 in Q20 bases; agarose-fp
Quality coverage: 6.01 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

*	1	1801: contig of 1801 bp in length
*	1802	1901: gap of unknown length
*	1902	4641: contig of 2740 bp in length
*	4642	4741: gap of unknown length
*	4742	9679: contig of 4938 bp in length
*	9680	9779: gap of unknown length
*	9780	13392: contig of 3613 bp in length

*	13493	17164: contig of 3672 bp in length
*	17165	17264: gap of unknown length
*	17265	21510: contig of 4246 bp in length
*	21511	21610: contig of unknown length

41011	27059: contig of 5479 bp in length
27190	27189: gap of unknown length
35636	contig of 8447 bp in length
35637	gap of unknown length

	contig	length
45538	gap of unknown length	10814
45638	contig of 10814 bp in length	
56452	gap of unknown length	10814
56551	gap of unknown length	10814

	gap of	length
66803	gap of	unknown length
66903	contig of	12216 bp in length
79119	contig of	unknown length
79219	contig of	15137 bp in length

	contig	length
94456	111274:	16819 bp
111275	111374:	gap of unknown length
111375	130600:	contig of 19226 bp
120601	120700:	gap of unknown length

	Location/Qualifiers	length
*	130701	150278: contig of 19578 bp in length
*	150279	150378: gap of unknown length
*	150379	179538: contig of 29160 bp in length

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"

```

```

nature
1.1801
/note="assembly_name:Contig13"
1902.4641
/note="assembly_name:Contig14"

```

```

          4742. .9079
          /note="assembly_name:Contig15"
          9780. .13392
          /note="assembly_name:Contig16"

```

```

/note="assembly_name:Contig17"
17265. .21510
/note="assembly_name:Contig18"
21511. 22000

```

```

/note="assembly_name:Contig19"
27190. . 35636
/note="assembly_name:Contig20"

```


misc_feature	35737. .45537	/note="assembly_name:Contig21"
misc_feature	45638. .56451	/note="assembly_name:Contig22"
misc_feature	56552. .66802	/note="assembly_name:Contig23"
misc_feature	66903. .79118	/note="assembly_name:Contig24"
misc_feature	79219. .94355	/note="assembly_name:Contig25"
misc_feature	94456. .111274	/note="assembly_name:Contig26"
misc_feature	11375. .130600	/note="assembly_name:Contig27"
misc_feature	130701. .150278	/note="assembly_name:Contig28"
	clone_end:SP6	vector_side:left"
misc_feature	150379. .179538	/note="assembly_name:Contig29"
BASE COUNT	42230 a 47023 c 46289 g 42392 t	1604 others
ORIGIN		

Query Match	25.1%	Score 116.4	DB 2	Length 179538
Best Local Similarity	65.6%	Pred. No. 9e-22		
Matches 231	Conservative	0	Mismatches 116	Indels 5
Gaps				4
QY 116	ccacggctgaacctacaagaccgcgaatctccgcgagaaattcagttglttctctctc	175		
Db 156227	CCGCCGCGAGCGTTGCAAAAGCTGGAAATCTCCGCGAGAAAGTCTCTTACTTCCACTTCC	156168		
QY 176	ccacattctcagggaaactcgaaaactcgcctctcg-gtgcagtcagatagcgct-g	233		
Db 156167	CACCTTCGCGGTGGCGGTGCTCGAAACCCGCGCTCTTCTCGAGATGACGACGCGGCGC	156108		
QY 234	cgtcagacataaactcccggtgtagtctglttgagcgagattgacttaagtgcga-gctt	292		
Db 156107	CGGTAGGCTATATAAACCTTAGTGCGCATATTTGTTGCTGCTTATGCGTTCCTCATCCCA	156048		
QY 293	gtgaaagattaatcgagaccccgcgagactcgcattccttltgcggagacagctltg	352		
Db 156047	GTTTCTTGATCTTATGCAAAAGCTGTGACAGACCCCGCCGCTTGTCCACGSGCACTTG	155988		
QY 353	cgacagcccgctgagacatacagt--cccccagccccacgcttgaaggcgacatgaacg	410		
Db 155987	AGGCAGCGCGGAGATACTCTGAGATTACTGCGAGACCCGAGCGCTGAGGGGAGATGAACGG	155928		
QY 411	ctgcgcttgaagagcaatccgcgacacagatcgcttlttgcaaacggacgg	462		
Db 155927	CTGCGCTTCCTTAACCGCTCGGACCTGTGATCCCTTCTGCGACCGGAACCGG	155876		
RESULT 6				
AC008749_3				
WPCOMMENT				
Sequence split into 5 fragments	LOCUS AC008749	Accession AC008749		
Fragment Name	Begin	End		
AC008749.0	1	110000		
AC008749.1	100001	210000		
AC008749.2	200001	310000		
AC008749.3	300001	410000		
AC008749.4	400001	467420		
Continuation (4 of 5) of AC008749	from base 300001	AC008749 Homo sapiens chro		
Query Match	24.7%	Score 114.8	DB 2	Length 110000
Best Local Similarity	65.3%	Pred. No. 2.6e-21		
Matches 230	Conservative	0	Mismatches 117	Indels 5
Gaps				4
QY 116	ccaaggtgacctcaaacgacccggaatctccggagaaattcagtggtgttctctctc	175		
Db 49833	CCGCGCGAGCGTTGCAAAAGCTGGAAATCTCCGCGAGAAAGTCTCTTACTTCCACTTCC	49892		

QY	176	caacattctcaaggagcttcgaaactcgcctctcgg	-gtgacgtcaagatagcgt	-g	233
Db	49893	CAACCTTCGGGTTTGGCGTGTCGAACACCCGCGCTCTCTTGCGATGCTACAGACCCCGGGG			49952
QY	234	cgccacagataaactccccggtgacgltgtlgtgcagatltgactcaagtcga	-gctt	292	
Db	49953	CGGTAGCGATATAAAGCCATAGGCGCATTTGGTCTTATGCTTATGCTTCCAAACCA		50012	
QY	293	gtgaaagattacatgcgaaaccgccgagatccgcacaccttgcggaagacgtcttg		352	
Db	50013	GTTGTTGATCTTATGTCAGACGCTGCACGCCCGCGCGCTTTTGCCACGGCACTTG		50072	
QY	353	cgacacgcccgtagacatcagt--ccccgaagcccaagccttgaggcgacatgaacgcg		410	
Db	50073	AGGCACGCGGAGTACTCTGACTTACTCGGAGCCCGAGCGCTGAGGGTGAATGAACGGG		50132	
QY	411	cttgagccttgagagcaatccgagaccagatcgcttttgccaacgaacgcg		462	
Db	50133	CTGGCCTCCCTAACCGTCCGGACCTGTGATCGCTTTTGCGACGACGAAACGG		50184	

RESULT	7
AC026803	AC026803 231450 bp DNA HTG 26-JUL-2001 Homo sapiens chromosome 19 clone CTD-2639E6, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
LOCUS	AC026803
DEFINITION	AC026803.5 GI:15022021
ACCESSION	AC026803.5
VERSION	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
KEYWORDS	human
SOURCE	Homo sapiens
ORGANISM	

REFERENCE	1 (bases 1 to 231450)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 231450)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	On Jul 26, 2001 this sequence version replaced g1:9256651.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

*	1	67040: contig of 67040 bp in length
*	67041	gap of unknown length
*	67141	231450: contig of 164310 bp in length.

Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 843459
Center clone name: CTD-2639E6

Summary Statistics

Consensus quality: 228571 bases at least Q40
Consensus quality: 230251 bases at least Q30
Consensus quality: 230412 bases at least Q20

```

Estimated insert size: 226000; agarose-fp estimation
Estimated insert size: 230499; sum-of-contigs
estimation
Quality coverage: 7.98 in Q20 bases; agarose-fp

```

```

      estimation      Quality coverage: 7.83 in Q20 bases; sum-of-contigs
      estimation.
      Location/Qualifiers
FEATURES             source
                      1..231450
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="19"
                      /clone="CTD-2639E6"
BASE COUNT          55608 a 59790 c 60166 g 55786 t 100 others
ORIGIN

Query Match
Best Local Similarity 24.7%; Score 114.8; DB 2; Length 231450;
Matches 230; Conservative 0; Mismatches 117; Indels 5; Gaps 4;

QY 116 ccaagcgagactcaagaagcccggaatcccgagagaattcagtggtttctctct 175
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 116775 CCGCGGTGACGTTGCAAAAGCTGGAATCTCCGCGAAGTCTGCTTACTTCCACTTCC 116834

QY 176 ccaccttctcaggagactccggaactccgcctctcg-gtgaactgaagcagtagcgtc-g 233
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116835 CACCTTTCGGGTTGCGGTCTGAAACCCGCCCTCTCTTCGTGACGTCAAGCAGCGGGGCG 116894

QY 234 cgtcaactataaactcccggtgtagctgtgtgagcagattgactcagtcgca-gctt 292
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116895 CCGTAGGCTATAAAGCCTAGTGGCCATTGTTGTTGCTTATGCTTATGCGTTCCATCCCA 116954

QY 293 gtggaagattacatgagagagcccgagcagatccgcatccctttgcccggagacgctttg 352
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116955 GTTGTATCTTATGCAAGAGCGTGCACGCCGCCGCCCTTGTGCGCCAGGCACTTGTG 117014

QY 353 cgaagcccggtgagacatcaagt--cccgagccccagccctgagagcgagacatgaacgcg 410
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117015 AGGACGCCGAGATACCTCTGAGTTACTCGAGAGCCCGAGCCTGAGGCTGAGATGAACCGC 117074

QY 411 ctggccttgagagcaatccgagaccagcatcgcttttggcaaacgagacgcg 462
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117075 CTGGCCTTCCTAACCCGTGCGACCTGTGATCGCTTGTGCGAGACCGAACC GG 117126

RESULT 8
LOCUS CRUGAD34X 2088 bp mRNA ROD 29-JUN-1994
DEFINITION Hamster Gadd34 mRNA, complete cds.
ACCESSION L28147
VERSION L28147.1 GI:452489
KEYWORDS acidic protein.
SOURCE Cricetus longicaudatus cDNA to mRNA.
ORGANISM Cricetus longicaudatus
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus.
REFERENCE 1 (bases 1 to 2088)
AUTHORS Zhan,Q., Lord,K.A., Alamo,I.Jr., Hollander,M.C., Carrier,F.,
TITLE Ron,D., Kohn,K.W., Hoffman,B., Liebermann,D.A. and Fornace,A.J.Jr.
The gadd and MYD genes define a novel set of mammalian genes
encoding acidic proteins that synergistically suppress cell growth
JOURNAL Mol. Cell. Biol. 14, 2361-2371 (1994)
MEDLINE 94187707
FEATURES             Location/Qualifiers
source              1..2088
                    /organism="Cricetus longicaudatus"
                    /db_xref="taxon:10030"
                    /cell_line="Chinese hamster ovary-K1"
                    215..1987
                    /gene="Gadd34"
                    215..1987
                    /gene="Gadd34"
                    215..1987
                    /note="Gadd34 is the hamster homolog of the mouse MYD16
                    cDNA; contains multiple PEST sequences implicated in rapid
                    protein turnover"

```

```

/codon_start=1
/protein_id="AAA36983.1"
/db_xref="gi:452490"
/translation="MAPSPROPHILWRDAHSFHLSPLMFLSRANSRLRPEAPPE
WPAETVVGADQIEADAPAPLVEHPPOGEAESEGTPEBKAQOCPDVOANSSP
PETGLSDDDKOGDGRREGORATAGLPLILSGLOSAKSLGEVAVAGEGVELAY
PTSHWEGCSPEEEDGFTVKAPRASDPSQKKSSTVYCPGEAEHATBEKOTENKA
DPSSPSGSRAMEYCSKOBGEADPEPHRAGKIQLOONAEAEBEBAKVSLSVSG
NAFLKAWYVRGEDTEDDDSDWSAESEKALSSPTSPEDFLKAWYVRGEDTEDD
DDSDWGSAAEEBEGKALSSPTSPEDFLKAWYVRGEDTEDDDSDWSAEKDLQTEA
TPTSAFLKTMVCCPGEDTEDDDDEVVYVPEDEADDPKSPSEAGOCPLPEGELEGY
EAESLFOVAFFYLPGEKPAPEWPAKPLPLRLORLTLRTPTODDEETPLRAKVF
SENVTVEFLAVMGAPQARRGPEWGLARBSRRARRIQAEBELGTYLTPARARAW
ARLGNPSLPLALEIDCHITFPFSQ"
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
polyA_signal
polyA_signal
polyA_signal
BASE COUNT          487 a 602 c 594 g 405 t
ORIGIN

```

```

Query Match
Best Local Similarity 24.6%; Score 114.2; DB 10; Length 2088;
Matches 142; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

QY 293 gtggaagattacatgagagcccgagcag-actcgaatccctttgcccggagacgcttt 351
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GTGGAAGATTGCATGCGAGACCCGCCGCCGCCGCCGATCCCTTCGTGCGGACCTCCT 60

QY 352 gcgacagcccggtgagacatcaagtcgccgagccccagccgctgagagcgagacatgaacgcg 411
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CGGCGCTGTGTGTGAGACCTCACGTCGCCGAGCCCGACGCTGAGGCGGACATGAACCTC 120

QY 412 tggccttgagagcaatccgagaccagcatcgcttttggcaaacgagacgcg 462
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 TGCCCTTGCGAGACCTCGGAGACCGGAGACCGCTTCTG6CACACGAAACCG 171

```

```

RESULT 9
LOCUS AC064877/c 165166 bp DNA HTG 29-JUN-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-782E3, WORKING DRAFT SEQUENCE,
2 unordered pieces.
ACCESSION AC064877
VERSION AC064877.5 GI:14575806
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165166)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165166)
AUTHORS Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (23-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 29, 2001 this sequence version replaced gi:14280310.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0782E03

```

```

----- Summary Statistics -----
Sequencing vector: MJ3; 498
Sequencing vector: plasmid; 508
Chemistry: Dye-terminator Big Dye; 53% of reads
Chemistry: Dye-terminator Big Dye; 53% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 164535 bases at least Q40
Consensus quality: 164782 bases at least Q30
Consensus quality: 164917 bases at least Q20
Insert size: 162000; agarose-ff
Insert size: 165066; sum-of-ctdigs
Quality coverage: 14.14 in Q20 bases; agarose-ff
Quality coverage: 13.88 in Q20 bases; sum-of-ctdigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 163910: contig of 163910 bp in length
* 163911 164010: gap of unknown length
* 164011 165166: contig of 1156 bp in length.
Location/Qualifiers
1. 165166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-782E3"
1. 163910
/note="assembly_name:Contig24"
164011.165166
/note="assembly_name:Contig4"
misc_feature
BASE COUNT 52669 a 30734 c 29916 g 51745 t 102 others
ORIGIN

Query Match 22.9%; Score 106.2; DB 2; Length 165166;
Best Local Similarity 61.2%; Pred. No. 5.9e-19;
Matches 237; Conservative 0; Mismatches 145; Indels 5; Gaps 4;

QY 72 gattacgagctctcgtgattgagcaagagcagcctgggaa-gccacggtgacctaac 130
|||||
Db 164397 GATTGACAGTTGCTTTGTTGAGGGGGTGTGACGCTCGAAGACTCCGCGTACGTTGC 164338
|||||

QY 131 aagggccggaatcctcgcggaagattcagtggtttctcctctccacattctcaagg 190
|||||
Db 164337 AAAAGCTGGAATCTCCGCGAGAGAGTCGTCTTACTTCCACTTCNNCCCTTCGGGTTGCG 164278
|||||

QY 191 acttcggaactcgcctctccggtgacgtcagcatagcgt-gcgtcagactataaact 249
|||||
Db 164277 GTCGTGAAACCCCGCTCTCTTCGTGACGTGACGCGCGGCGGAGGCTATAAAG 164218
|||||

QY 250 cccgggtgactggttggcgagattgactcagtcgca-gctgtggaagattacatgc 308
|||||
Db 164217 CCTAGGGCCATTGTTGCTGCTCTTATCCATCCATCCCATTTGTTGATCTTATGC 164158
|||||

QY 309 gagaccgagcagactcgcacatcccttgcggaagcagccttgcgaagcccgtagagc 368
|||||
Db 164157 AAGAGCTGACAGACCCCGCGCTGTCGCCACGCACTTGAGAGCGCGGAGATAC 164098
|||||

QY 369 atcagct--ccccgagccccagcctgagggagacatagaagcgtctgagagagcaa 426
|||||
Db 164097 TCTGAGTTACTCGAGCCCGACGCGCTGAGGAGTGAATGAACGCGCTCCCTCAACCG 164038
|||||

QY 427 tccgagccagcagcagccttggcgaagaa 453
|||||
Db 164037 TCCGAGACGTGTATCGCTTCTGGCACA 164011
|||||

```

```

RESULT 10
CGU83983 224 bp DNA ROD 07-JUL-1998
LOCUS CGU83983
DEFINITION Cricetus griseus apoptosis associated protein (GADD34) gene,
promoter sequence.
ACCESSION U83983
VERSION U83983.1 GI:3258620
KEYWORDS
SOURCE
ORGANISM
Chinese hamster.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus.
REFERENCE 1 (bases 1 to 224)
AUTHORS Hollander M.C., Zhan Q., Bae I. and Fornace A.J. Jr.
TITLE Mammalian GADD34, an apoptosis- and DNA damage-inducible gene
JOURNAL J. Biol. Chem. 272 (21), 13731-13737 (1997)
MEDLINE 97298078
REFERENCE 2 (bases 1 to 224)
AUTHORS Hollander M.C. and Fornace A.J. Jr.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD
20892, USA
FEATURES
source
Location/Qualifiers
1..224
/organism="Cricetus griseus"
/db_xref="taxon:10029"
<1..>224
/gene="GADD34"
/note="growth arrest and DNA damage-inducible gene 34"
<1..>224
/gene="GADD34"
GC_signal
135..146
/gene="GADD34"
149..158
/gene="GADD34"
misc_signal
174..179
/note="ATF/CRE site"
/gene="GADD34"
TATA_signal
BASE COUNT 46 a 67 c 63 g 48 t
ORIGIN

Query Match 20.7%; Score 96.2; DB 10; Length 224;
Best Local Similarity 73.5%; Pred. No. 6.3e-16;
Matches 164; Conservative 0; Mismatches 53; Indels 6; Gaps 3;

QY 67 tgcgagttgagcgtctctgattgagcgaagacagcctggga-aagcaggtgac 125
|||||
Db 3 TTCGAGATTGACACTCTGCANAGTTTGCGCGAGACAGCCGTGGAAAGTCGCAATGAC 62
|||||

QY 126 ctcaagcccggaatcctcgcggaattcagtggtttctcctctccacattctc 185
|||||
Db 63 CTCACAAAGCCCGGAATTCGCGAGAAATTGCGTGTGTTT--CCCTGCCACCTTTC 120
|||||

QY 186 cagggaactccgaactcgcctctccggtgacgtgagctagcagtcgctgagactata 245
|||||
Db 121 CCGGAGCTTCGGAAGCCCGCTCTCCGCTGTCGTCAGCA--GCTTCGCCAGGCTATA 177
|||||

QY 246 aactcccggtgacgtgattgagcagattgactcagtcga 288
|||||
Db 178 AAAGCAGTGTTCGGCTTCGCGCGCAGATTGAGCAGATTCCTCA 220
|||||

RESULT 11
HSU83981 2331 bp mRNA PRI 07-JUL-1998
LOCUS HSU83981
DEFINITION Homo sapiens apoptosis associated protein (GADD34) mRNA, complete
cDS
ACCESSION U83981
VERSION U83981.1 GI:3258617
KEYWORDS
SOURCE
human.

```

ARI00139	LOCUS	2344 bp	DNA	PAT	14-FEB-2001
ARI00139	DEFINITION	Sequence 2 from patent US 6080558.			
ARI00139	ACCESSION	ARI00139			
ARI00139	VERSION	ARI00139.1	GI:12810587		
ARI00139	KEYWORDS				
ARI00139	SOURCE	Unknown.			
ARI00139	ORGANISM	Unknown.			
ARI00139	REFERENCE	1 (bases 1 to 2344)			
ARI00139	AUTHORS	Bandman,O., Tal,P., Shah,P. and Corley,N.C.			
ARI00139	TITLE	Polynucleotide encoding human growth regulator protein			
ARI00139	JOURNAL	Patent: US 6080558-A 2 27-JUN-2000;			
ARI00139	FEATURES	location/Qualifiers			
ARI00139	source	1..2344			
ARI00139	BASE COUNT	556 a 667 c 709 g 412 t			
ARI00139	ORIGIN				
ARI00139	Query Match	17.5%; Score 81.2; DB 6; Length 2344;			
ARI00139	Best Local Similarity	70.9%; Fred. NO. 6.9e-12;			
ARI00139	Matches 122; Conservative	0; Mismatches 48; Indels 2; Gaps 1;			
ARI00139	Db	8 GTTGTTGATCTTATMGACAGCGTGCACAGACCCGCGCCGCTGTGTGCACAGGCACTTG 67			
ARI00139	Qy	293 gtgaaagattacatgcgagacccgcgcgcgacatcccttctgcgcgagacgccttg 352			
ARI00139	Db	353 cgaagcgcctgtgagctacagct--cccgagacccacgccttgagggagacatgaagcg 410			
ARI00139	Qy	68 AGGAGCGCGGAGACTCTGATGACTTACTGCGAGCCGACGCGCTGAGGGTGGATGAAGCGG 127			
ARI00139	Db	411 ctgcgccttgagagcaatccgagccacgacatcgctttggcaaacgcgacccg 462			
ARI00139	Qy	128 CTGCGCTCCTTAACCGTCCGAGACTGTGATCGCTTGTGGAGACCGAACCGG 179			
ARI00139	RESULT 13				
ARI00139	LOCUS	AK001361 2349 bp mRNA PRI 22-FEB-2000			
ARI00139	DEFINITION	Homo sapiens cDNA FLJ10499 f1s, clone NT2RP2000346, weakly similar to MEO1D0 DIFFERENTIATION PRIMARY RESPONSE PROTEIN MYD116.			
ARI00139	ACCESSION	AK001361			
ARI00139	VERSION	AK001361.1 GI:7022572			
ARI00139	KEYWORDS	oligo capping; f1s (full insert sequence).			
ARI00139	SOURCE	Homo sapiens testis/carcinoma cell line:NM2 cDNA to mRNA, clone:11b:NT2RP2 clone:NT2RP2000346.			
ARI00139	ORGANISM	Homo sapiens			
ARI00139	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
ARI00139	AUTHORS	1 (sites)			
ARI00139	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagaatsuma,M., Hosofiti,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,A., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuo,Y., Ninomiya,K. and Iwayanagi,T.				
ARI00139	TITLE	NEDO human cDNA sequencing project			
ARI00139	JOURNAL	Unpublished (2000)			
ARI00139	REFERENCE	2 (bases 1 to 2349)			
ARI00139	AUTHORS	Isogai,T. and Otsuki,T.			
ARI00139	TITLE	Direct Submission			
ARI00139	JOURNAL	Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)			
ARI00139	COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan. cDNA full insert sequencing: Research Association for Biotechnology. cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science,			

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE AUTHORS TITLE JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL	1 (bases 1 to 233) Hollander, M.C., Zhan, O., Bae, I. and Fornace, A.J. Jr. Mammalian GADD34, an apoptosis- and DNA damage-inducible gene J. Biol. Chem. 272 (21), 13731-13737 (1997) 97298078 2 (bases 1 to 233) Hollander, M.C. and Fornace, A.J. Jr. Direct Submission Submitted (07-JUN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD
REFERENCE AUTHORS TITLE JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL	1 (bases 1 to 233) Hollander, M.C., Zhan, O., Bae, I. and Fornace, A.J. Jr. Mammalian GADD34, an apoptosis- and DNA damage-inducible gene J. Biol. Chem. 272 (21), 13731-13737 (1997) 97298078 2 (bases 1 to 233) Hollander, M.C. and Fornace, A.J. Jr. Direct Submission Submitted (07-JUN-1997) LMP, NIH, Bldg. 37, Room 5C09, Bethesda, MD